

STIC-Biotech/ChemLib

117557

From: Kaushal, Sumesh
Sent: Tuesday, March 23, 2004 5:22 PM
To: STIC-Biotech/ChemLib
Subject: 09938391: Sequence and Interference Search

09/938,391 Sequence and Interference Search

Please search

SEQ ID NO:2 PRT 230

SEQ ID NO:4 PRT 184

SEQ ID NO:1 DNA 829

SEQ ID NO:3 DNA 555

thanks

S. Kaushal

AU1636, REM2.B85

Ph: 571-27-20769

Mail Box: REM2.C70

Interference Search files removed - 86

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 3/31/04
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

RECEIVED
MAR 24 2004
STIC

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 29, 2004, 09:11:19 ; Search time 3335.77 Seconds
(without alignments)
10771.549 Million cell updates/sec
Title: US-09-938-391-1
Perfect score: 829
Sequence: 1 cccgtggcggcagatgacat.....aataaagggaagcacaagag 829

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sw.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.tv.*
- 39: em.hgo.hum.*
- 40: em.hgo.mus.*
- 41: em.hgo.other.*

pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	829	100.0	829	6	AX399629 Sequence
2	829	100.0	829	6	BD188110 Method an
3	555	66.9	555	6	AX399631 Sequence
4	555	66.9	555	6	BD188111 Method an
5	550.4	66.4	552	6	BD240900 Expresio
6	504.8	60.9	2637	9	BC033715 Homo sapi
7	503.2	60.7	1388	9	AX130835 Homo sapi
8	503.2	60.7	5408	9	AF018082 Homo sapi
9	503.2	60.7	5929	9	AF018081 Homo sapi
10	501.6	60.5	2364	9	AK098216 Homo sapi
11	496.8	59.9	3394	6	I51045 Sequence 4
12	496.8	59.9	3394	6	AX409531 Sequence
13	496.8	59.9	3394	9	HUMCOL18AX
14	493.2	59.5	900	6	AX370851 Sequence
15	436.8	52.7	1615	10	MMU03714 Mus musculu
16	436.8	52.7	4437	10	MUSCOLLAG
17	436.8	52.7	4802	10	BC043697 Mus muscu
18	423	51.0	4031	6	I51044 Sequence 1
19	423	51.0	4031	10	MUSCOLPRO
20	420	50.7	564	9	AF416592 Homo sapi
21	418.4	50.5	563	6	AX431862 Sequence
22	418.2	50.4	2376	6	AX816106 Sequence
23	418	50.4	546	6	AX193165 Sequence
24	418	50.4	549	6	BD240882 Expresio
25	418	50.4	549	6	AX100086 Sequence
26	418	50.4	552	6	BD266741 Proteins
27	418	50.4	552	6	AX395662 Sequence
28	418	50.4	552	6	AX473835 Sequence
29	418	50.4	552	6	BD081407 Fused pro
30	418	50.4	786	9	AF282883 Homo sapi
31	418	50.4	2376	6	AX816108 Sequence
32	411.6	49.7	555	9	AF184060 Homo sapi
33	410.8	49.6	534	6	AX193166 Sequence
34	410.8	49.6	537	6	AX100092 Sequence
35	409.4	49.4	851	10	AF189709 Rattus no
36	398.4	48.1	650	6	AX370853 Sequence
37	394.2	47.6	552	6	BD272245 Anti-angi
38	377.8	45.6	624	6	AX084406 Sequence
39	376.8	45.5	573	6	BD271283 Compositi
40	376.8	45.5	573	6	AX183471 Sequence
41	376.8	45.5	573	6	AX183520 Sequence
42	376.8	45.5	573	6	AX217279 Sequence
43	376.8	45.5	573	6	AX282751 Sequence
44	376.8	45.5	573	6	AX042272 Sequence
45	376.8	45.5	573	6	AX490666 Sequence

ALIGNMENTS

RESULT 1	AX399629	829 bp	DNA	linear	PAT 06-JUN-2002
LOCUS	Sequence 1 from Patent EP1191036.				
DEFINITION	AX399629				
ACCESSION	AX399629				
VERSION	AX399629.1	GI:21335409			
KEYWORDS	Canis familiaris (dog)				
SOURCE	Canis familiaris				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
REFERENCE	1				
AUTHORS	Sheppard M.G. and Tong X.				
TITLE	Methods and compositions for diagnosing and treating disorders involving angiogenesis				

JOURNAL Patent: EP 1191036-A 1 27-MAR-2002;
Pfizer Products Inc. (US)
Location/Qualifiers
source 1. .829
/organism="Canis familiaris"
/mol_type="unassigned DNA"
/db_xref="taxon:9615"

ORIGIN

Query Match 100.0%; Score 829; DB 6; Length 829;
Best Local Similarity 100.0%; Pred. No. 1e-99;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTGGGGGACATGACATCTCTGGCGGCGCCCGCGCTCTGAGACCCCGCCCTAC 60
Db 1 CCTGGGGGACATGACATCTCTGGCGGCGCCCGCGCTCTGAGACCCCGCCCTAC 60

Qy 61 CCGGGGCGCCCGGACACACGCTCTCTAGTGCATCTCCAGCGGCTCGGCCACTGTTGG 120
Db 61 CCGGGGCGCCCGGACACGCTCTCTAGTGCATCTCCAGCGGCTCGGCCACTGTTGG 120

Qy 121 CCGGTCCAC 180
Db 121 CCGGTCCAC 180

Qy 181 CTGAACAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 181 CTGAACAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

Qy 241 CAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
Db 241 CAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300

Qy 301 CAGGACCTCTACAGCATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 301 CAGGACCTCTACAGCATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360

Qy 361 AGGACAGGCTCTCTCTCCCGAGTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGG 420
Db 361 AGGACAGGCTCTCTCTCCCGAGTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGG 420

Qy 421 AAGCCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 421 AAGCCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

Qy 481 CCGCGGAGAGCGGTGGCGAGCGGTCCGACCCAGCGGCGGCGGCGGCGGCGGCGGCG 540
Db 481 CCGCGGAGAGCGGTGGCGAGCGGTCCGACCCAGCGGCGGCGGCGGCGGCGGCGG 540

Qy 541 TGCAGACGTGGCGGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 600
Db 541 TGCAGACGTGGCGGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

Qy 601 GGCAGGCTGTGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 601 GGCAGGCTGTGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 660

Qy 661 GAGAACAGCGTCAATGCTCTCTTCCAGTAGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 661 GAGAACAGCGTCAATGCTCTCTTCCAGTAGGCGGCGGCGGCGGCGGCGGCGGCG 720

Qy 721 GAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 721 GAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780

Qy 781 CACGCTCACGTTTAAATGTAATCTCTCAAGAAATAAAGGAAGCAAGAG 829
Db 781 CACGCTCACGTTTAAATGTAATCTCTCAAGAAATAAAGGAAGCAAGAG 829

RESULT 2
BD188110
LOCUS 829 bp DNA linear PAT 17-JUL-2003

DEFINITION Method and composition for diagnosing and treating diseases relating to angiogenesis.

ACCESSION BD188110

VERSION BD188110.1 GI:32997849

KEYWORDS JP 2003000268-A/1.

SOURCE Canis sp.

ORGANISM Canis sp.

REFERENCE 1 (bases 1 to 829)
Sheppard, M.G. and Tong, X. / X/
Method and composition for diagnosing and treating diseases relating to angiogenesis

TITLE Patent: JP 2003000268-A 1 07-JAN-2003;
PFIZER PRODUCTS INC

JOURNAL OS Canis sp. (dog)
PN JP 2003000268-A/1
PD 07-JAN-2003
PF 24-AUG-2001 JP 2001254697
PR 25-AUG-2000 US 60/227924
PI MICHAEL GEORGE SHEPPARD, XIAO TONG
PC C12N15/09, A01K67/027, A61K31/7089, A61K35/76, A61K39/395 PC
A61K39/395, A61K45/00,
PC A61K48/00, A61P7/00, A61P9/00, A61P17/02, A61P17/06, A61P17/08, A61P19/08, A61P25/00, A61P27/02, A61P27/06, A61P29/00,
PC A61P31/00, A61P35/00, A61P35/02, A61P35/04, A61P37/06, A61P43/00,
PC C07K14/47,
PC C07K16/18, C07K19/00, C12N1/15, C12N1/19, C12N1/21, C12N5/10 PC
C12Q1/02, G01N33/15,
PC G01N33/50, C12N15/00, C12N5/00
CC CANINE PRO ENDOSTATIN NUCLEOTIDE SEQUENCE

FT Key Location/Qualifiers
1. .829
FT source /organism="Canis sp. (dog)".
FT Location/Qualifiers
1. .829
/organism="Canis sp."
/mol_type="genomic DNA"
/db_xref="taxon:9615"

ORIGIN

Query Match 100.0%; Score 829; DB 6; Length 829;
Best Local Similarity 100.0%; Pred. No. 1e-99; Mismatches 0; Indels 0; Gaps 0;
Matches 829; Conservative 0;

Qy 1 CCTGGCGGCGAGATGACATCTCTGGCGGCGCCCGCGCTCTGAGACCCCGCCCTAC 60
Db 1 CCTGGCGGCGAGATGACATCTCTGGCGGCGCCCGCGCTCTGAGACCCCGCCCTAC 60

Qy 61 CCGGGGCGCCCGGACACACGCTCTCTAGTGCATCTCCAGCGGCTCGGCCACTGTTGG 120
Db 61 CCGGGGCGCCCGGACACACGCTCTCTAGTGCATCTCCAGCGGCTCGGCCACTGTTGG 120

Qy 121 CCGGTCCAC 180
Db 121 CCGGTCCAC 180

Qy 181 CTGAACAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 181 CTGAACAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

Qy 241 CAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
Db 241 CAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300

Qy 301 CAGGACCTCTACAGCATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 301 CAGGACCTCTACAGCATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360

Qy 361 AGGACAGGCTCTCTCTCCCGAGTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGG 420
Db 361 AGGACAGGCTCTCTCTCCCGAGTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGG 420

Db 361 AGGGACGAGGTGCTCTTCCAGCTGGAGGCTTATTCTCGGGCTCGAGGGCCAGCTG 420
QY 421 AAGCCCGGGCCCGCATCTTCTTTTCAGCGCAGAGATGCTTCAGCAGACCCCGCCTGG 480
Db 421 AAGCCCGGGCCCGCATCTTCTTTTCAGCGCAGAGATGCTTCAGCAGACCCCGCCTGG 480
QY 481 CCCCAGAAAGAGCGTGTGGCACTGCTTCAGCGCAGAGATGCTTCAGCAGACCTAC 540
Db 481 CCCCAGAAAGAGCGTGTGGCACTGCTTCAGCGCAGAGATGCTTCAGCAGACCTAC 540
QY 541 TCGAGACGTGGCGAGCGAGCGCCCGCGGCGCCAGCGGCGGCGGCGGCGGCGGCGG 600
Db 541 TCGAGACGTGGCGAGCGAGCGCCCGCGGCGCCAGCGGCGGCGGCGGCGGCGGCGG 600
QY 601 GGCAGGCTCTCGAGCAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 601 GGCAGGCTCTCGAGCAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
QY 661 GAGAACAGCGTGTATGTAATCTTCCAGTAAGTGGCGCGGCGGCGGCGGCGGCGG 720
Db 661 GAGAACAGCGTGTATGTAATCTTCCAGTAAGTGGCGCGGCGGCGGCGGCGGCGG 720
QY 721 GAGGGGGCGCGCGAGGAGCATCCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGG 780
Db 721 GAGGGGGCGCGCGAGGAGCATCCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGG 780
QY 781 CACCGTCAGCTTATGTAATCTTCCAGTAAGTGGCGCGGCGGCGGCGGCGGCGG 829
Db 781 CACCGTCAGCTTATGTAATCTTCCAGTAAGTGGCGCGGCGGCGGCGGCGGCGG 829

RESULT 3
AX399631
LOCUS 555 bp DNA linear PAT 06-JUN-2002
DEFINITION Sequence 3 from Patent EP1191036.
ACCESSION AX399631
VERSION AX399631.1 GI:21335410
KEYWORDS
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris (dog)
REFERENCE 1
AUTHORS Shepherd, M.G. and Tong, X.
TITLE Methods and compositions for diagnosing and treating disorders involving angiogenesis
JOURNAL Patent: EP 1191036-A 3 27-MAR-2002;
Pfizer Products Inc. (US)
FEATURES
source
1. .555
/organism="Canis familiaris"
/mol_type="unassigned DNA"
/db_xref="taxon:9615"

ORIGIN
Query Match 66.9%; Score 555; DB 6; Length 555;
Best Local Similarity 100.0%; P-red. No. 1e-63;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CACACCCAGGAGCTTCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 198
Db 1 CACACCCAGGAGCTTCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 60
QY 199 GCGGCGATGCGAGGATCCCGGCGAGCGGAGCTTCCAGTGGTGGTGGTGGTGGTGGTGG 258
Db 61 GCGGCGATGCGAGGATCCCGGCGAGCGGAGCTTCCAGTGGTGGTGGTGGTGGTGGTGG 120
QY 259 GGGTGGCGGCGACCTTCCGCGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318
Db 121 GGGTGGCGGCGACCTTCCGCGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 319 GTGCGCGCGCGCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 378

Db 181 GTGCG 240
QY 379 CCCAGCTGGAGGGCTTATTCTCGGGCTCCAGGGCCAGCTGAAGCCCGGGCCCGCATC 438
Db 241 CCCAGCTGGAGGGCTTATTCTCGGGCTCCAGGGCCAGCTGAAGCCCGGGCCCGCATC 300
QY 439 TTCTCTTTTCAGCGCAGAGATGCTCTGCGAGCAGCCCGCTGGCGGCGGCGGCGGCGG 498
Db 301 TTCTCTTTTCAGCGCAGAGATGCTCTGCGAGCAGCCCGCTGGCGGCGGCGGCGGCGG 360
QY 499 CACGGCTCCGACCCCG 558
Db 361 CACGGCTCCGACCCCG 420
QY 559 GAGGCG 618
Db 421 GAGGCG 480
QY 619 GAGGCG 678
Db 481 GAGGCG 540
QY 679 TCCTTCTCCAGTAG 693
Db 541 TCCTTCTCCAGTAG 555

RESULT 4
BD188111
LOCUS 555 bp DNA linear PAT 17-JUL-2003
DEFINITION Method and composition for diagnosing and treating diseases relating to angiogenesis.
ACCESSION BD188111
VERSION BD188111.1 GI:32997850
KEYWORDS JP 2003000268-A/2.
SOURCE Canis sp.
ORGANISM Canis sp.
REFERENCE 1 (bases 1 to 555)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
TITLE Shepherd, M.G. and Tong, X.
METHOD and composition for diagnosing and treating diseases relating to angiogenesis
JOURNAL Patent: JP 2003000268-A 2 07-JAN-2003;
Pfizer Products Inc
COMMENT OS Canis sp. (dog)
PN JP 2003000268-A/2
PD 07-JAN-2003
PF 24-AUG-2001 JP 2001254697
PR 25-AUG-2000 US 60/227924
PI MICHAEL GEORGE SHEPPARD, XIAO TONG
PC C12N15/09, A01K67/027, A61K31/7088, A61K35/76, A61K39/395 PC
A61K39/395, A61K45/00,
PC A61K48/00, A61P7/00, A61P9/00, A61P15/00, A61P17/00, PC
A61P17/02,
PC A61P17/06, A61P19/02, A61P19/08, A61P25/00, A61P27/02, A61P27/06,
PC A61P29/00,
PC A61P31/00, A61P35/00, A61P35/02, A61P35/04, A61P37/06, A61P43/00,
PC C07K14/47,
PC C07K16/18, C07K19/00, C12N1/15, C12N1/19, C12N1/21, C12N5/10 PC
C1201/02, G01N33/15,
PC G01N33/50, C12N15/00, C12N5/00
CC CANINE ENDOSTATIN NUCLEOTIDE SEQUENCE
FH Key Location/Qualifiers
FT source 1. .555
/organism="Canis sp. (dog)".
FT Location/Qualifiers
1. .555
/organism="Canis sp."
/mol_type="genomic DNA"
/db_xref="taxon:9616"

FEATURES
source
1. .555
/organism="Canis sp."
/mol_type="genomic DNA"
/db_xref="taxon:9616"

ORIGIN

Query Match 66.9%; Score 555; DB 6; Length 555;
Best Local Similarity 100.0%; Pred. No. 1e-63;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CACACCCACAGGACTTCAGCTGGTGTGACCTGTGGCCCTGAACAGCCCGGCGG 198
DB 1 CACACCCACAGGACTTCAGCTGGTGTGACCTGTGGCCCTGAACAGCCCGGCGG 60

QY 199 GGGCGCATCGAGGATCCCGGGAGCGGACTTCAGTGTCTTCCAGAGGCGCGCGCG 258
DB 61 GGGCGCATCGAGGATCCCGGGAGCGGACTTCAGTGTCTTCCAGAGGCGCGCGCG 120

QY 259 GGGCTGGCGGACCTTCCTGGGCTTCCTGTGCTGGGCTGAGGACCTCTACAGATC 318
DB 121 GGGCTGGCGGACCTTCCTGGGCTTCCTGTGCTGGGCTGAGGACCTCTACAGATC 180

QY 319 GTGCGCGCGCGGACCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 378
DB 181 GTGCGCGCGCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 240

QY 379 CCAGCTGGAGGCTTATTCCTGGGCTTCAGGAGCGGAGCTTCAGTGTCTTCCAGATC 438
DB 241 CCAGCTGGAGGCTTATTCCTGGGCTTCAGGAGCGGAGCTTCAGTGTCTTCCAGATC 300

QY 439 TTCTCTTTTCGAGGAGAGATGTCTGAGCACCCTGGGCTGGGCGGAGAGCGTGTGG 498
DB 301 TTCTCTTTTCGAGGAGAGATGTCTGAGCACCCTGGGCTGGGCGGAGAGCGTGTGG 360

QY 499 CAGGCTTCGAGCGGCGGCGGCTGAGGAGCGGAGCTTCAGTGTCTTCCAGATC 558
DB 361 CAGGCTTCGAGCGGCGGCGGCTGAGGAGCGGAGCTTCAGTGTCTTCCAGATC 420

QY 559 GAGGCGCGGCGGCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 618
DB 421 GAGGCGCGGCGGCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 480

QY 619 GAGGCGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 678
DB 481 GAGGCGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 540

QY 679 TCCTTCTCCAAG 693
DB 541 TCCTTCTCCAAG 555

RESULT 5
LOCUS BD240900 552 bp DNA linear PAT 17-JUL-2003
DEFINITION Expression and export of angiogenesis inhibitors as immunofusins.
ACCESSION BD240900
VERSION BD240900.1 GI:33050670
KEYWORDS JP 2002523036-A/20.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 552)
AUTHORS Lo, K.M., Li, Y. and Gillies, S.D.
TITLE Expression and export of angiogenesis inhibitors as immunofusins
JOURNAL Patent: JP 2002523036-A 20 30-JUL-2002;
LEXIGEN PHARMACEUTICALS CORP
COMMENT OS Canis familiaris (dog)
PN JP 2002523036-A/20
PD 30-JUL-2002
PF 25-AUG-1999 JP 2000566305
PR 25-AUG-1998 US 60/097883
PI KIN MING LO, YUE LI, STEPHEN D GILLIES
PC C12N15/09, A61K31/711, A61K38/00, A61K39/395, A61K48/00, A61P7/00, A61P35/00.
PC A61P43/00, C07K14/745, C07K14/78, C07K19/00, C12N5/10, C12P21/02, C12N15/00,
PC C12N5/00, A61K37/02
CC Endostatin

FEATURES
source 1..552
/organism="Canis familiaris"
/mol_type="genomic DNA"
/db_xref="taxon:9615"

ORIGIN
Query Match 66.4%; Score 550.4; DB 6; Length 552;
Best Local Similarity 99.8%; Pred. No. 4.1e-63;
Matches 551; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 139 CACACCCACAGGACTTCAGCTGGTGTGACCTGTGGCCCTGAACAGCCCGGCGG 198
DB 1 CACACCCACAGGACTTCAGCTGGTGTGACCTGTGGCCCTGAACAGCCCGGCGG 60

QY 199 GGGCGCATCGAGGATCCCGGGAGCGGACTTCAGTGTCTTCCAGAGGCGCGCGCG 258
DB 61 GGGCGCATCGAGGATCCCGGGAGCGGACTTCAGTGTCTTCCAGAGGCGCGCGCG 120

QY 259 GGGCTGGCGGACCTTCCTGGGCTTCCTGTGCTGGGCTGAGGACCTCTACAGATC 318
DB 121 GGGCTGGCGGACCTTCCTGGGCTTCCTGTGCTGGGCTGAGGACCTCTACAGATC 180

QY 319 GTGCGCGCGCGGACCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 378
DB 181 GTGCGCGCGCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 240

QY 379 CCAGCTGGAGGCTTATTCCTGGGCTTCAGGAGCGGAGCTTCAGTGTCTTCCAGATC 438
DB 241 CCAGCTGGAGGCTTATTCCTGGGCTTCAGGAGCGGAGCTTCAGTGTCTTCCAGATC 300

QY 439 TTCTCTTTTCGAGGAGAGATGTCTGAGCACCCTGGGCTGGGCGGAGAGCGTGTGG 498
DB 301 TTCTCTTTTCGAGGAGAGATGTCTGAGCACCCTGGGCTGGGCGGAGAGCGTGTGG 360

QY 499 CAGGCTTCGAGCGGCGGCGGCTGAGGAGCGGAGCTTCAGTGTCTTCCAGATC 558
DB 361 CAGGCTTCGAGCGGCGGCGGCTGAGGAGCGGAGCTTCAGTGTCTTCCAGATC 420

QY 559 GAGGCGCGGCGGCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 618
DB 421 GAGGCGCGGCGGCGGAGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 480

QY 619 GAGGCGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 678
DB 481 GAGGCGCGGAGCTTCAGTGTCTTCCAGAGGCGCGCGCG 540

QY 679 TCCTTCTCCAAG 690
DB 541 TCCTTCTCCAAG 552

RESULT 6
LOCUS BC033715 2637 bp mRNA linear PRI 08-JUL-2002
DEFINITION Homo sapiens, similar to collagen, type XVIII, alpha 1, clone IMAGE:4425380, mRNA, partial cds.
ACCESSION BC033715
VERSION BC033715.1 GI:21708148
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2637)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2390, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA library Preparation: Life Technologies
 CDNA library Arrayed by: The I.M.A.G.E. Consortium
 DNA Sequencing by: National Institutes of Health
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nhi.nih.gov
 Akter, N., Ayele, K., Beckstrom-Sternberg, S.I.,
 Blackley, R.W., Bouffard, G.G., Green, K., Br,
 Diatchek, N.L., Granite, S., Guan, X., Gupta, J.,
 Hansen, N., Ho, S.-L., Karlins, E., Kuong, P., J.,
 Maduro, Q.L., Masello, C., Maskrip, B., Mastr,
 McDowell, J., Pearson, R., Stratirop, S., Thon,
 Tsurgou, C., Vogt, J.L., Walker, M.A., Wether,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK plate: 68 Row: e Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES source

```

location/Qualifiers
1. .2637
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4425380"
/tissue_type="kidney, hypernephroma"
/clone_lib="NTH-MCC-99"
/lab_host="DH1.0B"
/note="vector: pCMV-SPORT6"

```

CDS

```

/codon_start=3
/protein_id="Stmilar to collagen, type XVIII, alpha 1"
/db_xref="GI:21708149"
/translation="TRPPRGRRGPPFCLCPGCPGREGPPGRTQKGLSGEAGP
GHWKSKGAPGAGRGESGLAGAPGPPGPPGPPGPPGLADWDMSGSGPPW
STAKSADGPPGLPGLKXGDPVGLPLGAKGEVADGAPGLPGREGIAGAPQPK
GDRGSKGDKGDPGVGGLPFGPPGPPGVVYSEGDGSLVSGPPGEGPPGAPG
GAPGAKGSLGKGRGSPGKGEKPPGPIFSGDGLAGSKGAKGEGPPGPPGPPG
GRPYGEGTIGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
YDYSNVFASRRGPPGLPFGNPGPPGKAGEVGPFPGLFLEAEKMKK
GDRGAKGKEGEGGGGFGSSLPFGPPGPPGPPGPPGPPGPPGPPGPPGPPG
GPGIYGEGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
LQPPVQLGQVHEVPEGHWLIIVAEQGLVYVQGNPKVQLEARTPLPRTGNVAA
WQTPVQMLDQNGPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPP
PAKPTSPQAHSHRDFQVLHLVALNPSAGSGMRGADFQCFQQAAGVLACTFRF
LSRLRGLQYIVARRADAAVPIVNLKDELLFSPMEALSPGKFKGAFIFPDQK
VHRYIVPICQNSVWMSGDNGRLTETSYCTWRTETAPSATGQSSLLGRLLGQAASC
HRAVTVICIKNSVFWATK"

```

ORIGIN

Query Match	60.9%;	Score 504.8;	DB 9;	Length 2637;
Best Local Similarity	77.7%;	Pred. No. 2.4e-57;		
Matches 665;	Conservative 0;	Mismatches 152;	Indels 39;	Gaps 3;

y	1	CCCTGGGGGCGACATGACATCTGCGCGGCCCCCGGCGCTGCTGAGACCCCGAGCCCTAC	60
b	1776	CCCTGGGGGCGACATGACATCTGCGCGAGCCCCCTCGCTGCCGAGCCCGAGCCCTAC	1835
y	61	CCCGGGGCCCGGACACACGGCTCTCTACGTGACATTCACAGCCGCTCGCCCCACTGGTGGG	120
b	1836	CCCGAGCCCCCGACACACAGCTCTCTACGTGACCTGGGCGCGCGGACCCACAAG----	1891
y	121	CCCGTCCACACCCACACCCACACCCAGGACATTCGAGTGTGTGCTGTGACTGTGTGCC	180
y	1892	-----CCACCCGCGACACCGGACATTCAGCGCGTGTCTCAACTGTGTGCG	1943

QY	181	CTGACACCGCCGACCGCGCGCGATCGCAGAGCATCTCCGGAGCGGACCTCCAGTGCCTTC	2403		
Db	1944	CTCAACAGCCCCCTGTCTGGCGGCGCATCGCGGCGCATCCCGGGCGCATCTTCAGTGCCTTC	2003		
QY	241	CAGCAGCGCGCGCGCGCGCGGCTGGCGCGACCTTCCTCGGGGCTTCCTCTGCTCGCGGCTG	300		
Db	2004	CAGCAGCGCGCGCGCGCGCGGCTGGCGGCGCATCTTCCTCGCGCGCTTCCTCTGCTCGCGCTG	2063		
QY	301	CAGGACCTCTACAGCATCGTGGCGCGCGCGGACCGCAGCGGGGTCCCGTCTGTCACACTC	360		
Db	2064	CAGGACCTCTACAGCATCGTGGCGCGCGCGGACCGGCGCGGCGCATCTGTCACACTC	2123		
QY	361	AGGACGAGGTGCTCTTCCCGACGCTGGGAGGCTTATTTCTGGGGTTCGAGGGCGAGCTG	420		
Db	2124	AAGGACGAGCTGCTGTCTTCGACGCTGGGAGGCTCTCTGCTCAGGGCTCTGAGGGTCCGCTG	2183		
QY	421	AAGCCCGGGCGCGCATCTCTTTTCGACGCGCAGAGATGTCCTTCAGCACACCCCGCTGG	480		
Db	2184	AAGCCCGGGCGCGCATCTCTTTCTTTGACGCGAGAGAGCTCTTCGAGCACCCACCTGG	2243		
QY	481	CCCCCGAAGAGCGTGTGGCACGCTCCGACCCCGAGCGCGCGCGCTGACCGACGACTAC	540		
Db	2244	CCCCAAGAGCGTGTGGCATGCTCGGACCCCGAGCGCGCGCATGACCGAGAGCTAC	2303		
QY	541	TGCGAGAGCTGCGGACGAGAGCGCGCGCGCGCGCGCATCGCGGCGAGGCTGCTCGCTGCTGG	600		
Db	2304	TGTGAGAGCTGCGGACGAGAGGCTCTCTCGGCGCACGGGCGAGGCTCTCTGCTGCTGGG	2363		
QY	601	GGCAGGCTGCTTGAGCAGGAGGCGCGAGTGCCTCGCGCACGCGCTTCGTTGGTCTCTGCACT	660		
Db	2364	GGCAGGCTCTTGCGGCGAGAGTGCCTCGCGCATGCGCTACATGCTGCTGCTGCACT	2423		
QY	661	GAGAACAGCTCATGACTCTCTTCTCAAGTAGGCGCGCGCGCGCGCGCATGCGGCGAGG	720		
Db	2424	GAGAACAGCTCTCATGACTGCTCTCAAGTAGCACCAGCGCTGAGTGGAGTGGCGGAGG	2483		
QY	721	GAGGCGGCGCGCGCAGGAGCATCGCGCGCGCGCGG-----	756		
Db	2484	ACCGCGGCTCGGAGGAGCG	2543		
QY	757	--GGCGCTGGCGGAGCGTTGCTGCAAGC--TCAGTTTATGTAATCTCAGAAATA	813		
Db	2544	CAGGACCTTGGCTGCGACTTCTCTATAGTTTACGTTTCTATGTAATCTCAGAAATA	2603		
QY	814	AAAGGAAGCCAAAGAG	829		
Db	2604	AAAGGAAGCCAAAGAG	2619		
RESULT 7	AK130835	1388 bp	mRNA	linear	PRI 10-SEP-2003
LOCUS	AK130835	Homo sapiens	cdNA FLJ27325	fis	clone TWS08226, highly similar to
DEFINITION	AK130835	Collagen alpha 1(XVII)	chain precursor.		
ACCESSION	AK130835				
VERSION	AK130835.1	Gi:34527728			
KEYWORDS	cligo capping; fis	(full insert sequence).			
SOURCE	Homo sapiens	(human)			
ORGANISM	Homo sapiens				
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					
2 (bases 1 to 1388)					
Sugano,S. and Suzuki,Y.					
Direct Submission					
Submitted (31-JUL-2003)	Sumio Sugano, Institute of Medical Science,				
Unpublished					

Direct Submission
Submitted (31-JUL-2003) Sumio Sugano, Institute of Medical Science,

COMMENT

NEEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction and 5'-end one pass sequencing: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; 3'-end one pass sequencing: RAB; clone selection for full insert sequencing: RAB and Helix Research Institute.

FEATURES

```
1. 1388
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="TMS08226"
/tissue_type="thymus"
/clone_lib="TMS"
/notes="cloning vector. ph
```

```
ORIGIN
/Note="Cloning vector: pMET8SFL3"
```

Query Match 60.7%; Score 503.2; DB 9; Length 1388;
Best Local Similarity 77.6%; Pred. No. 4.8e-57;
Matches 664; Conservative 0; Mismatches 153; Indels 39; Gaps 3;

1	CCCTGGGGCGGACGATGACATCTCTGGCCGGCCCGCCCGCGCTGCTGAGACCCCGACGCTAC	60
545	CCCTGGGGCGGACGATGACATCTCTGGCCAGCCCGCCCTCGCTCTCCGAGACCCCGACGCTAC	604
61	CCGGGGCCCGGACCAACAGCTCTCTACGTGCACTTCCAGCCGGGTGCGCCCACTGTGGG	120
605	CCCGAGCCCGCACCAACAGCTCTCTACGTGCACTTCCAGCCGGCCGGCCGACCCACAAG---	660
121	CCGCTCCACACCCACACCCACACACCCACAGGACTTCCAGCTGTGCTGCTGACCTGGTGCC	180
661	-----CCACCCGCCACAGCACCCCGACTTCCAGCCGGTGTCTCCACTGGTTCG	712
181	CTGAACAGCCCGAGCCGGCGGGCATGCGAGGCATCCCGGGAGCGGACTTCCAGTGTTC	240
713	CTCAACAGCCCGCTGTACGGCGGCATGCGGGGCATCCCGGGGCCGACTTCCAGTGTTC	772
241	CAGCAGCGCGCCCGCGGGGTGGCCGCACTTCCGGGCTTCTCTGTGTCGTCCGCGCTG	300
773	CAGCAGCGCGGCGCGTGGGGCTGGCGGCACCTTCCGGCTTCTCTGTCTCGGCGCTG	832
301	CAGGACTCTACAGATCTGTGGCGCGCGCCGACGCCACCGGGGTGCCGCTGCTCAACCTC	360
833	CAGGACTCTACAGCATCTGTGGCGCGTGGCCGACCGCGAGCCGTGCCCATCTGTCAACCTC	892
361	AGGGACGAGTGTCTTTCGCCAGCTGGGAGGCTTATTCTCGGGCTCCGAGGGCCAGCTG	420
893	AAGGACGAGTGTGTCTTTCAGCTGGGAGGCTCTGTCTCAGGCTCTGAGGCTCGGCTG	952
421	AAGCCGGGGCCCGCATCTTCTTTCGACGCGAGATGTCCTGACGACCCCGCTGG	480
953	AAGCCGGGGCACCGCATCTTCTTTGACGACGAGGACGCTCTGAGGACCCCACTCG	1012
481	CCCGGAAGAGGTGTGGACCGCTCCGACCCAGCCGGGCGCCCTGACCGACGAGCTAC	540
1013	CCCGAAGAGCGTGTGGCATGSGCTCGGACCCCAACGGGGCGAGGCTGACCGAGAGCTAC	1072
541	TGCGAGACGTGGCGGACGAGGCCCGCGGCCACCGGGCAGGCGGTGCTGCTGCTGGCG	600
1073	TGTGAGACGTGGCGGACGAGGCTCCCTCGGCCACGGGCAAGGCTCTCGCTGCTGGG	1132
601	GGCAGCTGCTGAGCAGAGGACCGGAGCTGCCGCCAGGCTTCTGTGTGCTCTGCACTC	660
1133	GGCAGGCTCTTGGGGCAGAGTCCCGGAGCTGCCATACGGCTACATCTGTGCTCTGCATT	1192
661	GAGAACAGCGTCTATGACCTCTTCTTCCAAGTAGGGCCCGCGGGGCCACGACAGGGGGG	720
1193	GAGAACAGTCTTATGACTGCTTCCAAGTAGCCACCGCTGATGCGATGCGATGCGCGAGG	1252

ORIGIN		Query Match	Score	DB	Length
RPTSPPAHSHRDFQPVLLHLVALNSPLSGMGRIGADPQCQOQARVAGLAGTFRAPFL		60.7%	503.2	DB 9	5408
SRLODLYSVTRADRAAYPIVNLKDELLFPSEWALFSGSEGPLKPGARIFISFDGKDLV		Best Local Similarity	77.6%	Pred. No. 3.2e-57	
RHTWPQKSWTGSBPNGRLLTESYCEWTRTEADPSATQOASLLGGRLLGQAASACH		Mismatches	0	Indels	39
AYIVLCIENSPFTASK"		Conservative	664	Gaps	3
1	CCCTGGCGGCAGATGACATCTCTGGCGCGCCCGCCCGCCCTGCTGGAGCCGCCAGCCCTAC	60			
3393	CCCTGGCGGCAGATGACATCTCTGGCGCGCCCGCCCGCCCTGCTGGAGCCGCCAGCCCTAC	3452			
61	CCCGGGCCCCGACACGAGCTCTACGPGCACTTCCAGCCGGCTGCGCCGCACTGTGTGGG	120			
3453	CCCGAGCCCGACACAGCTCTACGTCACCTTCAGCTGCGCGCGCGCGCACCAAG----	3508			
121	CCCGTCCACACCCACACCCACACACAGGACTTCCAGCTGCTGTCACCTGTGTGGCC	180			
3309	-----CCACCCGCCACAGCCACCGGACTTCCAGCCGGTGTCTCACCTGTGTGG	3560			
181	CTGAACAGCCCGCAGCGGGCGGCATGCGAGGCATTCGGGGAGCGGACTTCCAGTGTTC	240			
3561	CTCAAGAGCCCTGTTCAGCGGCATGCGGGGCATTCGCGGGCGGACTTCCAGTGTTC	3620			
241	CAGCAGCGCGCGCGCGGGCTGCGCGGCACCTTCCGGGCTTCTCTGTGTGTGTGTGTGT	300			
3621	CAGCAGCGCGCGCGGGCTGCGGGGCATCTTCCGGGCATCTTCTGTCTCTGTGTGTGTGT	3680			
301	CAGGACCTCTACAGCATCTGTGCGCGCGCGCACCGACCGGGGTGCCGTGTGTAACCTC	360			
3681	CAGGACCTGTACAGCATCTGTGCGCGCTGCGCACCGCGAGCGGTGCCATCGTCAACCTC	3740			
361	AGGACAGAGTGTCTTCCCGAGCTGGGAGGCTTATCTCGGGCTCCGAGGGCCAGCTG	420			
3741	AAGACAGAGTGTGTGTTCAGCTGGGAGGCTGTGTCTCAGGCTGTGAGGTTCCGCTG	3800			
421	AAGCCCGGGCGCGCATCTTCTTTTCGACGGCAGAGATGCTCTGCAGCACCCCGCTGG	480			
3801	AAGCCCGGGCACGCATCTTCTTTGACGCCAAGGACGCTCTGAGCACCCACCTGG	3860			
481	CCCGGAGAGCTGTGTGGACGCTCCGACCCAGCGGGCGCGCTGACCGACAGCTAC	540			
3861	CCCGAGAGAGCGTGTGGCATGTGCTCGGACCCCAACGGGCGCGGCTGACCGAGAGCTAC	3920			
541	TGCGAGAGCTGGCGAGCGAGGCGCGCGCGCACCGGCGAGGCTGTCTGCTCTGGCG	600			
3921	TGTGAGAGCTGGCGGACGAGGCTCCTCGGACACGGGCGAGGCTCTCTGCTGTGGG	3980			
601	GGCAGGCTGTGGAGCAGGAGCGCGAGTGGCGGCACGCTTGTGTGTGTGTGTGTGTGT	660			
3981	GGCAGGCTCTTGGGCGCAGAGTGGCGGAGCTGCCATCAGCCCTACATCGTCTGTGATT	4040			
661	GAGAACGCTCATGACTCTTCTTCCAGTAGGGCGCGCGCGCACCGGACAGCGCGGGG	720			
4041	GAGAACGCTCATGACTCTTCTTCCAGTAGGCTGCTTCAAGTAGCCACCGCTGTGATG	4100			
721	GAGGGGCGCGCGCAGGAGCATCGCGCGCGCGGG-----	756			
4101	ACCGGCGCTCGGAGGAAGCCCGACCGTGGGCGAGGAGCGCGCGCGCCCTGTGCC	4160			
757	--GGGCTTGGCGGAGCGCTGTGCTGCACCG--TCAGTTTATGTAACTCTCAAGAAATA	813			
4161	CAGGACCTGGCTGCCATCTTCTGTATAGTTTACGTTTCAATGTATCTTCAAGAAATA	4220			
814	AAAGGAGCCAAAGAG	829			
4221	AAAGGAGCCAAAGAG	4236			

AF01:8081 5929 bp mRNA linear PRI 18-MAR-1999
 Homo sapiens type XVIII collagen (COL18A1) mRNA, alternatively
 spliced, long form, complete cds.
 AF01:8081
 AF01:8081.1 GI:2920534
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 5929)
 Saarela,J., Ylikarppa,R., Rehn,M., Purmonen,S. and Pihlajanieni,T.
 Complete primary structure of two variant forms of human type XVIII
 collagen and tissue-specific differences in the expression of the
 corresponding transcripts
 Matrix Biol. 16 (6), 319-328 (1998)
 98164096
 9503365
 2 (bases 1 to 5929)
 Saarela,J., Ylikarppa,R., Rehn,M., Purmonen,S. and Pihlajanieni,T.
 Direct Submission
 Submitted (28-JUN-1997) Dept. of Medical Biochemistry, University
 of Oulu, Kajaniemie 52 A, Oulu 90220, Finland
 Location/Qualifiers
 1..5929
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1..5929
 /gene="COL18A1"
 41..4591
 /gene="COL18A1"
 /note="alternatively spliced; long form (NC1-493)"
 /codon_start=1
 /product="type XVIII collagen"
 /protein_id="AAC39658.1"
 /db_xref="GI:2920535"
 /translation="WAPVPGCGHILLFLCCLAAARALLNLNLWLFNNEDTSHAATT
 IPEQPLVQQTATTHVTRPNSSTEPATPGSPPESELLEDGQDPTTGAESPDA
 PENTAGVAGRLNVAKGRIYRFLVNDWTPTESLARAEITVLETPESGLVAPGSSIT
 PEGNTLTPWSPGSGPQATTEAGTLFAPTPSPSPGRWPAPITGCVSPVPPSPERI
 SEVGLLQLLPPQVQVQTTDDPGLVAFVFGPDANSQVARYHPFSLFFRDFSLFL
 HIRPATGEGVLFITDSAAVMVLGLSVQDGHODISLITPFGAQTHTAASFR
 LPAFVQWTHLALSIVAGGVALVYDCEPQRMPLARSRGLEPGAGLFTVAQAGAD
 PKFGQVIAELKVRDPQVSPMHCIDLEDGDDSGAFGDSGSLGDAERLREITGAAL
 KRLPAPPVPTVPLAGSSTEDSRSEVEQTTVASLGAOTLPQSDSVSTWDGVRP
 PGRVVRGKCKGKEGPGVPGPRAGAPGSGCLPQPGPLPCPVSPPLGAPGAKDQVGE
 RQDGPQPGRDGTGDRDEGDPGDEGKPTGQPGFTGPDGVPKGDGKDGDPGVGE
 RGPQPGQPGPSPGSPFGRDKLFTIDMEGSGFGLLEALRPRGPGPPGPPGPGVGLP
 GPGPRGVNSDVPGPAGLPGVPGVGGPGLPGPPGPPGPPGPPGPPGPPGPPGPPG
 AGAPHGKSGAPGAPAGESGLAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 GPPWTSADGPGPPGLGKLGPGVPLGKAGEVADGTPGPPGPPGPPGPPGPPGPPGPPG
 QPKDGRSGRKGPPGPPGLGKLGPGVPLGKAGEVADGTPGPPGPPGPPGPPGPPGPPG
 ADFPGAPGKNGLSKGRSGSPGPKGKEPSSISPDGAGLGPAGKAGKGGPPGPPG
 PPGPYGKYGITGPPGRPGRMNGLKGEKPGPDASLGPPGPPGPPGPPGPPGPPGPPG
 PCTPVYDYNFAESRPPGPPGLPGNQPGPKAGEVGGPPGPPGPPGPPGPPGPPGPPG
 KGEKDGKDGAGKGEKGEKGGGFGSSLPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 KQGPQPGPTGYEGRQPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 SGVRLWATQMLQVGHVEPREGWLIIVAEQELIVRVQNGFRKVLQEARTPILPRGTDN
 EYVALQPPVQLHDSNPPEBHEPHTREMRADDILASPRRLPEPQVPYGPAPHHSY
 VHLRPARPSPFAHSRDPQVHLVALNSPLSGMGEGIRGADFQCFQOARAVGLAGT
 FRAFLSSRLQDYSIVRRADRAAPVILNKLDFLSEWALFGSGSGPLKFGARIISF
 DGDVLRHTPTQKSVNMGSDPNRRLTSEYCEITWATEAPSTQASSLGRLLQQS
 AASCHAYIVLCIENSFMTASK"

Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuoka, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, S., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2364)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FUJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomic@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'- end one pass sequencing: RAS,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAS; annotation: HRI and RAS.
Location/Qualifiers
1. .2364
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UTER02002964"
/tissue_type="uterus"
/clone_lib="UTERU2"
/note="cloning vector: DME18SFL3"

Query Match	60.5%	Score 501.6	DB 9	Length 2364
Best Local Similarity	77.5%	Prod. No. 6.7e-57		
Matches 663	Conservative	0	Mismatches 154	Indels 39
Gaps	39			
QY	1	CCCTGGCGGCAGATCATCTCTGGCCGGCCCGCCCTCTCTGTGACCCCGCCAGCCCTAC	60	
DB	1518	CCCTGGCGGCAGATGACATCTCTGGCCCGCCCGCCCTCTCTGTGACCCCGCCAGCCCTAC	1577	
QY	61	CCCGGGCCCGCACACAGCGCTCTTACGTGCATCTCAGCCGCTCTCCGCCACTGTGGG	120	
DB	1578	CCCGGAGCCCGCACACAGCTCTTACGTGCATCTGCGGCTGTGCACACCCACAAG----	1633	
QY	121	CCCTCCACACCCACACCCACAGGACTTCCAGCTGTGTCTGTGCACTCTGTGGCC	180	
DB	1634	-----CCACCCCGCCACAGCCACCCGACCTCCAGCCGGTCTCCACCTGTTGG	1685	
QY	191	CTGAAGACCCCGCAGCCGGCGGCGATGCGAGGCATCCGGGAGCGGACTTCCAGTGCCTC	240	
DB	1686	CTAACAGCCCCCTGTCTAGCGCGGCATGCGGGGCATTCGCGGGCGGACTTCCAGTGCCTC	1745	
QY	241	CAGCAGCGCGCGCCGCGGGGTGTGCGCGGCACCTTCCTGGGCTTCTGTCTGTGCGGCTG	300	
DB	1746	CAGCAGCGCGCGCCGTGGGGCTGGCGGGACCTTCCGCGCTTCTCTCTCTCGGCTG	1805	
QY	301	CAGGACCTCTACAGCATCTGTGGCGCGCGCGACCGGCACCGGGGTGCGCGTCAACCTC	360	
DB	1806	CAGGACCTGTACAGCATCTGTGGCGCTGCCACCGCGAGCCGTGCGCATCTGTCAACCTC	1865	
QY	361	AGGAGCAGAGTGTCTTCCGCCAGCTGGAGGACCTTATCTCGGCTCCGAGGGCCAGCTG	420	
DB	1866	AAGGACAGTGTGTCTTCCAGCTGGAGGCTCTGTCTCAGGCTCTGAGGTCCTCGCTG	1925	
QY	421	AAGCCCGGGCCGCACTTCTCTTTCAGCGGAGATGTCCTGACGACCCCGGCTCG	480	
DB	1926	AAGCCCGGGGCACGCATCTTCTCTTTTACCGGCAAGGACGTCTCTGAGGACCCCACTCG	1985	
QY	481	CCCGGAAGAGCGTGTGGCAGCGGCTCCAGCCCGGGGCGCGCTGTACCGGACGACTAC	540	
DB	1986	CCCGAAGAGCGTGTGGCATGCTCTCGACCCCAACGGGCGAGGCTGACCGGAGACTAC	2045	

Query Match	60.5%	Score 501.6	DB 9	Length 2364
Best Local Similarity	77.5%	Prod. No. 6.7e-57		
Matches 663	Conservative	0	Mismatches 154	Indels 39
Gaps	39			
QY	1	CCCTGGCGGCAGATCATCTCTGGCCGGCCCGCCCTCTCTGTGACCCCGCCAGCCCTAC	60	
DB	1518	CCCTGGCGGCAGATGACATCTCTGGCCCGCCCGCCCTCTCTGTGACCCCGCCAGCCCTAC	1577	
QY	61	CCCGGGCCCCGACACACAGCGCTCTTACGTGCATCTCAGCCCGCTCCGCCACTGTGGG	120	
DB	1578	CCCGGAGCCCCGACACACAGCTCTTACGTGCACCTGGCGCTGCACACCCACAAG----	1633	
QY	121	CCCTTCACACCCACACCCACAGCACTTCCAGCTGTGTCTGTGCACCTGTGTGGCC	180	
DB	1634	-----CCACCCCGCCACAGCCACCCGACCTTCCAGCCGGTCTCCACCTGTTGG	1685	
QY	191	CTGAAGACGCCGACGCGGGCGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCCTTC	240	
DB	1686	CTAACAGCCCCCTGTCTAGGGCGGCATGCGGGGCATTCGCGGGCGGACTTCCAGTGCCTTC	1745	
QY	241	CAGCAGCGCGCGCCGCGGGGTGGCGCGCACTTCCGGGCTTCTGTGTCTCGCGGCTG	300	
DB	1746	CAGCAGCGCGGSCCGTGGGGCTGGCGGGCACTTCCGGCGCTTCTGTCTCTCGCGCTG	1805	
QY	301	CAGGACCTCTACAGCATCTGTGGCGCGCGCGACCGCACCGGGGTGCGCGTCAACCTC	360	
DB	1806	CAGGACCTGTACAGCATCTGTGGCGCGTCCGACCGCGAGCCGTGCCCATCTGTCAACCTC	1865	
QY	361	AGGAGCAGAGTGCTCTTCCGCCAGCTGGAGGACCTTATCTCGGCTCCGAGGGCCAGCTG	420	
DB	1866	AAGGACGAGTGCTGTTTCCAGCTGGAGGCTCTGTCTTCAGGCTCTGAGGTCCTCGCTG	1925	
QY	421	AAGCCCGGGCCGCACTTCTCTTTCAGCGGAGATGTCCTGACGACCCCGGCTCG	480	
DB	1926	AAGCCCGGGGCACGCATCTTCTCTTTTACCGGCAAGGACGTCTCTGAGGACCCCACTGG	1985	
QY	481	CCCGGAAGAGCGTGTGGCACGGCTCCGACCCCGGGGCGCGCTGTGACCGACACTAC	540	
DB	1986	CCCGAAGAGCGTGTGGCATGCTCTCGACCCCAACGGGCGAGGCTGACCGAGAGCTAC	2045	

541 TCGGAGAGTGGCGACGAGGCCCCCGGCGCACCGGGCAGGGCTGTCGCTGCTGGCG 600
2046 TGTGAGAGTGGCGACGAGAGGCTCCCTCGGCCACGGGCCAGGCTCTCTCGCTGCTGGGG 2105
601 GCGAGGCTGCTGGAGCAGAGGCGCGGAGCTGCGCGCACCGCTTCGTGGTCTGCAATC 660
2106 GCGAGGCTCTGGGCGACAGTGCAGGAGCTGCGGAGCTGCCATCAGCGCTACATCGTGTCTGCAAT 2165
661 GAGACAGGTCATGACCTCTCTCCAGAGTAGGGCGCGCGGCCACGACGAGGCGGG 720
2166 GAGAACAGCTTCATGACCTCTCCAGTAGCCACCGCTCCAGTGCAGATGCGCGGAGAGG 2225
721 GAGGGGGCGCGCGCAGGAGCATCGCGCCCGCCCGGG----- 756
2226 ACCGCGGCTCGGAGGAGCGCCACCGTGGGAGGAGCAGCGCGCAGCGCCCTGGCCC 2285
757 --GGGCTTGGCGGAGCGCTTGGCTGCGCACCG--TCAAGTTTAATGTAATCTCAAGAAATA 813
2286 CAGGACCTGGCTGCGCATACTTCTCTGTATAGTTCACGTTTCATGTAATCTCAAGAAATA 2345
814 AAAGGAAGCCAAAGAG 829
2346 AAAGGAAGCCAAAGAG 2361
RESULT 11
LOCUS I51045 3394 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 4 from patent US 5643783.
ACCESSION I51045
VERSION I51045.1 GI:2472748
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3394)
AUTHORS Olsen,B.R. and Oh,S.P.
TITLE Collagen and uses therefor
JOURNAL Patent: US 5643783-A 4 01-JUL-1997;
FEATURES
source
1..3394
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 59.9%; Score 496.8; DB 6; Length 3394;
Best Local Similarity 77.1%; Pred. No. 2.5e-56;
Matches 660; Conservative 0; Mismatches 157; Indels 39; Gaps 3;
1 CCTGGCGGCGAGATGACATCTCTGCGCGGCCCGCGCGCTGCTGGACCCCGCCCTAC 60
1378 CCTGGCGGCGAGATGACATCTCTGCGCGGCCCGCGCGCTGCGCGAGCCCGCCCTAC 1437
61 CCGGGGGCGCGCACCGAGCTCTACGTGCACTTCCAGCGCGGCTCGCGCCCACTGGTGG 120
1438 CCGGAGGCGCGCACCGAGCTCTACGTGCACTGCGCGCGCGCGCGCGAGCCACCAAG--- 1493
121 CCGGTCCACACCGACACCGACACCGAGCTTCCAGCTGGTGTGCTGCACTGCTGGCGC 180
1494 -----CCACACCGCGCGACCGCGAGCTTCCAGCGCGGCTGCTCCACTGGTGG 1545
181 CTGAACAGCGCGCGCGCGCGATGCGAGCATCCCGGAGAGCGACTTCCAGTGTCTC 240
1546 CTCACAGCGCGCGCGCGCGCGATGCGGCGCATCCCGGCGCGCGCGCGCGCGCGCTC 1605
241 CAGCAGCG 300
1606 CAGCAGCG 1665
301 CAGGACCTTACAGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
1666 CAGGACCTGTACAGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1725

361 AGGACAGAGGTGCTCTTCCCGCAGCTGGGAGCGCTTATTTCTCGGGCTCCGAGGGCCAGCTG 420
1726 AAGACAGAGTGTCTTCCCGCAGCTGGGAGCGCTTGTCTCAGGCTCTGAGGGTCCGCTG 1785
421 AAGCCCGGGCGCGCGCGCTCTTCTTTCGAGCGGAGAGATGTCCTGCGAGCACCGCGCTGG 480
1786 AAGCCCGGGCGCGCGCGCTCTTCTTTCGAGCGGAGAGATGTCCTGAGGCGACCCCGCTGG 1845
481 CCCCGAAGACGCTGTGGCAGCGCTCCGAGCCCGCAGCGGGCGCGCGCTGACCGACAGCTAC 540
1846 CCCCGAAGACGCTGTGGCAGCGCTCCGAGCCCGCAGCGGGCGCGCGCTGACCGAGAGCTAC 1905
541 TGCAGAGCTGGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGGCG 600
1906 TGTGAGAGCTGGCGGAGGAGCGCTCCCTCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGGGG 1965
601 GCGAGGCTGTGGGAGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTGTGCTGCTG 660
1966 GCGAGGCTTCTGGGCGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCTTCTGTGCTGCTGCTG 2025
661 GAGAACAGCGTCATGACCTCTTCTTCCAGTAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGGG 720
2026 GAGAACAGCTTTCATGACTGCTCCCAAGTAGCCACCGCTGATGTCAGATGGCGCGAGAGG 2085
721 GAGGGGGCGCGCGGAGGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG--- 756
2086 ACCGCGGCTCGGAGGAGCG 2145
757 --GGGCTTGGCGGAGAGCTTGGCTGCGCACCG--TCAAGTTTAATGTAATCTCAAGAAATA 813
2146 CAGGACCTGGCTGCCATATCTTCTGTATAGTTCACGTTTCATGTAATCTCAAGAAATA 2205
814 AAAGGAAGCCAAAGAG 829
2206 AAAGGAAGCCAAAGAG 2221
RESULT 12
LOCUS AX409531 3394 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2178 from Patent WO0229103.
ACCESSION AX409531
VERSION AX409531.1 GI:21442236
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Alves,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2178 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source
1..3394
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. L22548"
ORIGIN
Query Match 59.9%; Score 496.8; DB 6; Length 3394;
Best Local Similarity 77.1%; Pred. No. 2.5e-56;
Matches 660; Conservative 0; Mismatches 157; Indels 39; Gaps 3;
1 CCTGGCGGCGAGATGACATCTGCGCGGCCCGCGCGCTGCTGGACCCCGCCCTAC 60
1378 CCTGGCGGCGAGATGACATCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1437
61 CCGGGGGCGCGCACCGAGCTCTACGTGCACTTCCAGCTGGTGTGCTGCACTGCTGGCGC 120
1438 CCGGAGGCGCGCACCGAGCTCTACGTGCACTGCGCGCGCGCGCGCGAGCCACCAAG--- 1493
121 CCGGTCCACACCGACACCGAGCTTCCAGCTGGTGTGCTGCACTGCTGGCGC 180
1494 -----CCACACCGCGCGACCGCGAGCTTCCAGCGCGGCTGCTCCACTGGTGG 1545
181 CTGAACAGCGCGCGCGCGCGATGCGAGCATCCCGGAGAGCGACTTCCAGTGTCTC 240
1546 CTCACAGCGCGCGCGCGCGCGATGCGGCGCATCCCGGCGCGCGCGCGCGCGCGCTC 1605
241 CAGCAGCG 300
1606 CAGCAGCG 1665
301 CAGGACCTTACAGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
1666 CAGGACCTGTACAGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1725

121	CCGCTCACACCACACCCACACCCACAGGACTTCAGCTGTGTGTGCACTGTGTGGC	180
1494	-----CCACCCCGCCACAGCCACCCGACTTCCAGCGGTGTCTCCACTGTGTGG	1545
181	CTGAACAGCCCGCAGCGCGGGCATGAGAGCATCCGGGAGCGGACTTCCAGTGCCTTC	240
1546	CTCAACAGCCCTGTTCAGCGGCATGCGGGCATCCCGGGGCCGACTTCCAGTGCCTTC	1605
241	CAGCAGCGCGCGCGCGGGCTGTGGCCGCGCACTTTCGGGCTTCTGTGTGTGTGGGCTG	300
1606	CAGCAGCGCGGGCGCGTGGGGCTGTGGCGGCACTTTCGGCGCTTCTGTGTGTCTCGCGCTG	1665
301	CAGGACCTCTTACAGACATCGTGTGGCGCGCGCACCGGAGTGCCTGTGTGTGTGTGTGTGT	360
1666	CAGGACCTGTACAGCATCGTGTGGCGTGTGGCGACCGCGCAGCGCGTCCATCGTCAACCTC	1725
361	AGGAGCAGATGTCTTTCACAGCTGTGGAGGCTTATTCTGGGCTCCGAGGCGCAGCTG	420
1726	AAGGACAGCTGTGTGTTCAGAGCTGGGAGGCTCTGTCTCAGGCTCTGAGGGTGTGGCTG	1785
421	AAGCCCGGGGCGCGCATCTTCTCTTTCGACGCGCAGAGATGTCTCTGACACCCCGGCTGG	480
1786	AAGCCCGGGGCAAGCATCTTCTCTTTCAGCGCAAGGACGCTCTGAGGCAACCCACTGG	1845
481	CCCAGGAGAGCTGTGTGGCAGGCTTCGACCCACGCGGGCGCGCTTGACGACACTAC	540
1846	CCCAGAGAGCGGTGTGGCATGTGCTCGGACCCCAACGCGGCGCAGGCTGACCGAGAGCTAC	1905
541	TGCAGAGCTGTGGCGAGCGAGGCGCCCGCGGCCACCGCGGCAGGGCTGTGTGTGTGTGGG	600
1906	TGTGAGAGCTGTGGCGAGCGAGGCTCCCTCGGCCACGGGCGAGGCTCTCTGTGTGTGGG	1965
601	GGCAGGCTGTGTGAGCAGGAGCGCGGAGCTGTGCGGCACGCTTGTGTGTGTGTGTGTGTGT	660
1966	GGCAGGCTCTTGGGGCAGAGTGTGCGGAGCTGCCATACGCTCATCATGTGTGTGTGTGTGT	2025
661	GAGAACAGCGTCATGACTCTTCTCTCAAGTAGGGCGCGGGGCCACGACGACGAGCGGGG	720
2026	GAGAACAGCTTCATGACTGTGCTTCAAGTAGCCACCGCTGTGATGACAGTGTGCGGAGG	2085
721	GAGGGGGCGCCGAGGAGCATTCGGCGCCCGGGG-----	756
2086	ACCGCGGCTCGGAGGAAGCCCGCACCGTGTGGCAGGGAGCAGCGCGGCGAGCCCTGTGCC	2145
757	--GGGCTTGGCGGAGCGCTTGT	813
2146	CAGGACCTGGTGTGCCATCTTCTCTGTATAGTTCAGTTCATGTATCTTCAAGAAATA	2205
814	AAAGGAAGCCAAAGAG	829
2206	AAAGGAAGCCAAAGAG	2221

RESULT 13	
HUMCOL18AX	
LOCUS	3394 bp mRNA linear PRI 01-NOV-1994
DEFINITION	Human collagen type XVIII alpha 1 (COL18A1) mRNA, partial cds.
ACCESSION	L22548
VERSION	L22548.1 GI:348908
KEYWORDS	alpha-1 type XVIII collagen.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 3394)
AUTHORS	Oh, S.F., Warman,M.L., Seldin,M.F., Cheng,S.D., Knoll,J.H., Timmons,S. and Olsen,B.R.
TITLE	Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the alpha 1(XVIII) collagen gene to mouse chromosome 10 and human chromosome 21
JOURNAL	Genomics 19 (3), 494-499 (1994)
MEDLINE	94245237
PUBMED	8188291

[illegible]

ORIGIN

Query Match	59.9%;	Score 496.8;	DB 9;	Length 3394;
Best Local Similarity	77.1%;	Pred. No. 2.5e-56;		
Matches 660;	Conservative	0;	Mismatches 157;	Indels 39; Gaps 3;
QY	1	CCCTGCGGGCAGATGACATCTGGCCGCCGCCGCCGCGCTGTGTGACGCCCCAGCCCTAC	60	
DB	1378	CCCTGCGGGCAGATGACATCTGGCCGCCGCCGCCGCGCTGTGTGACGCCCCAGCCCTAC	1437	
QY	61	CCCGGGGCCCGCACCAACCGCTCTACTGTGCACTTCCAGCCGGCTGCCCCACTGTGGG	120	
DB	1438	CCCGAGGGCCCGCACCAACCGCTCTACTGTGCACTTGGCGGCCGCGACAGCCCAAG----	1493	
QY	121	CCCGTCCACACCCACACACCCACACGACTTCCAGCTGGTGTGTGACCTGTGTGGCC	180	
DB	1494	-----CCACCCGCCACACGACCCGCGACTTCCAGCCGGTGTCTCACCTGTGTGG	1545	
QY	181	CTGAACAGCCCGCAGCGCGCGGCATCGAGGCATCCGGGGAGCGACTTCCAGTGCTTC	240	
DB	1546	CTCAACAGCCCCGTGTACAGGGGCAATCGGGGCATCCGGGGCCGACTTCCAGTGCTTC	1605	
QY	241	CAGCAGCGCGCGCGCGGGCTGGCGGGCACTTTCGGGCCCTTCTGTGTGTGCGGGCTG	300	
DB	1606	CAGCAGCGCGCGCGCGGGCTGGCGGGCACCTTTCGGGCCCTTCTGTGTGTGCGGGCTG	1665	
QY	301	CAGGACTCTTACAGCATCTGTGCGCGCGCGACCGCACCGGGGTGCCCGTGTGTCAACCTC	360	
DB	1666	CAGGACTGTGTACAGCATCTGTGCGCGCGGTGCCGACCGCGCAGCGTGTGCCACTGTGT	1725	
QY	361	AGGGACAGAGTGTCTTCTCCACGCTGGGAGCCCTTATTTCTGGGCTCCGAGGCGCAGCTG	420	
DB	1726	AAGGACAGAGTGTCTTCTCCACGCTGGGAGCTCTGTCTTCTCAGGGCTGTGAGGGTCCGCTG	1785	
QY	421	AAGCCCGGGGCCCGCATCTTCTCTTTCGACGGGAGAGATGTCTCTGACGACACCCGCTGG	480	
DB	1786	AAGCCCGGGGCCCGCATCTTCTCTTGTGCGGAAGGACGTCTCTGTAGGACACCCACTGG	1845	
QY	481	CCCGGGAAGAGCGTGTGGCACCGGCTCCGACCCGAGGGGGCGCGCTGTGACCGACAGCTAC	540	
DB	1846	CCCAAGAGAGCGTGTGGCATGTGCTCGACCCCAACGGGGCGAGGCTGTGACGAGAGCTAC	1905	
QY	541	TGCGAGACGTGGCGGACGAGAGGCCCGCGGCGGCCACCGGGCAGGCGTCTGTGTGTGGCG	600	

Db 1906 TGTGACGCTGGCGGACGAGGCTCCCTCGGCCACCGGCCAGGCTCCTCGCTGCTGGG 1965
QY 601 GGCAGGCTCTGGAGCAGAGAGCGGAGCTGCCCGACGCTTCGTTGGTCTCTGCATC 660
Db 1966 GGCAGGCTCCTGGGCGAGAGTGC CGGAGCTGCCATCAGCTCATCATGCTCTGCATT 2025
QY 661 GAGAACAGGCTCATGACCTTCCTTCCAAAGTAGGCGCGCGGCCACGACAGCGCGGG 720
Db 2026 GAGAACAGCTTCATGACTCCCTCCAAAGTAGCACCGCTCGATGCGCGGAGG 2085
QY 721 GAGGGGGCGCGCAGAGCATCGCGCGCCCGGG----- 756
Db 2086 ACCGGCGGCTCGGAGGAAGCCCCACCGTGGCAGGAGCAGCGCGCGCGCTGGCCCC 2145
QY 757 --GGGCTCGCGGAGCGCTGCTGCACCG-TCACGTTTAAATGTAATCTCAAGAAATA 813
Db 2146 CAGGACTGGCTGCCATATTTCTGTATAGTTACGTTTCTATGTAATCTCAAGAAATA 2205
QY 814 AAAGGAAGCCAAAGAG 829
Db 2206 AAAGGAAGCCAAAGAG 2221

RESULT 14
AX370851
LOCUS AX370851 900 bp DNA linear PAT 01-MAR-2002
DEFINITION Sequence 12 from Patent WO0210372.
ACCESSION AX370851
VERSION AX370851.1 GI:19168980
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE 1
AUTHORS Chapman, P.W., de Luca, G. and Falcicola, L.
TITLE Method of producing functional protein domains
JOURNAL Patent: WO 0210372-A 12 07-FEB-2002;
Applied Research Systems ARS Holding S.A. (AN)
LOCATION/Qualifiers
FEATURES
source 1..900
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="mRNA for mGSP fused to COL18A1 exons 38-41 for pEnd-HR#1"

ORIGIN
Query Match 59.5%; Score 493.2; DB 6; Length 900;
Best Local Similarity 83.9%; Pred. No. 1.1e-55;
Matches 574; Conservative 0; Mismatches 98; Indels 12; Gaps 1;
QY 1 CCCTGGCGGCGAGATGACATCTCTGGCGGACCCCGCGGCTCTGACCCCGCCCTAC 60
Db 211 CCCTGGCGGCGAGATGACATCTCTGGCGAGCCCGCTCGCTCGCGAGCCCGCCCTAC 270
QY 61 CCCGGGCGCGGACACGCTCTCTAGTGCATTCACGCGGCTCGCCCGCCACTGGTGG 120
Db 271 CCGGAGCCCGGACACGCTCTCTAGTGCATTCGCGCGGCGGACCCACAAAG--- 326
QY 121 CCGTCCACACCCACACCCACACCCACGAGACTTCACGCTGCTGCTGACCTGCGGCC 180
Db 327 -----CCACCCCGCCACAGCACCCGCGACTTCACGCGGCTGCTCACCTGGTGG 378
QY 181 CTGAACACCGCGAGCGGCGGCTGCGAGGATCCGGGAGCGGACATTCAGTGGCTTC 240
Db 379 CTCACAGCCCGCTGTGAGGGGATGCGGGGATGCGGGGCGGACATTCAGTGGCTTC 438
QY 241 CAGCAGGCGGCGCGCGGCGGCTGGCGGACATTCGCGGCGCTTCGTTGCTGCGGCTG 300
Db 439 CAGCAGGCGGCGGCGGCTGGCGGCTGGCGGACATTCGCGGCGCTTCGTTGCTGCGGCTG 498
QY 301 CAGGACCTCTACGATCTGCGGCGGCGGCGGACCGGACCGGCGGCTGCTGCTCAACCTC 360

Db 499 CAGGACCTGTATCAGCATCTGTCGCGCTGCCGACCGCGCAGCGCTGTCCTCAACCTC 558
QY 361 AGGACGAGGCTGCTTCCAGCTGGGAGGCTTATTTCTCGGCTCGAGGGCCAGCTG 420
Db 559 AAGACGAGCTGCTGTTTCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGCTCCGCTG 618
QY 421 AAGCCCGGGCGCGCATCTTCTCTTTCCAGCGCAGAGTCTCTGACGACACCCGCGCTGG 480
Db 619 AAGCCCGGGCGCACGATCTTCTCTTTGACGCGCAAGGACGCTCTGAGGCACCCACCTGG 678
QY 481 CCGCGGAGAGGCTGTGACGCGCTCCGACCCCGCGGCGCGCTGACGACGAGCTAC 540
Db 679 CCGCGAGAGAGGCTGTGCGATGGCTCGGACCCCGCGGCGGAGCTGACGAGGCTAC 738
QY 541 TCGGAGAGCTGCGGACGAGAGCGCGCGCGCGCGCGCGGCGGAGGCTGCTGCTGCTGGCG 600
Db 739 TGTGAGAGCTGCGGACGAGGCTCCCTCGGCGACGCGGCGGCGGCTCTGCTGCTGGGG 798
QY 601 GCGAGCTCTGAGCAGGAGCGCGGAGCTGCCGCGGCTGCGCGCGCTTCTGCTGCTGCAATC 660
Db 799 GCGAGCTCTGCGGCGAGGTCGCGGAGCTGCCATCAGGCTACGCTGCTGCTGCTGCAAT 858
QY 661 GAGAACAGCGCTCATGACCTCTTC 684
Db 859 GAGAACAGCTTCATGACTGCTTC 882

RESULT 15
MMU03714
LOCUS MMU03714 1615 bp mRNA linear ROD 13-AUG-1994
DEFINITION Mus musculus BALB/c alpha 1 type XVIII collagen (COL18A1) mRNA,
partial cds.
ACCESSION U03714
VERSION U03714.1 GI:487733
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Rehn, M.V. and Pihlajaniemi, T.
TITLE A previously uncharacterized collagen chain, alpha 1(XVIII), with
frequent interruptions in the collagenous sequence, a distinct
tissue distribution, and homology with type XV collagen
Proc. Natl. Acad. Sci. U.S.A. (1994) In press
JOURNAL 2 (bases 1 to 1615)
REFERENCE Rehn, M., Hintikka, E. and Pihlajaniemi, T.
AUTHORS Primary structure of the alpha 1 chain of mouse type XVIII
TITLE collagen, partial structure of the corresponding gene, and
comparison of the alpha 1(XVIII) chain with its homologue, the
alpha 1(XV) collagen chain
J. Biol. Chem. 269 (19), 13929-13935 (1994)
JOURNAL 94245707
MEDLINE 818673
REFERENCE 3 (bases 1 to 1615)
AUTHORS Rehn, M.V.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1993) Marko Rehn, Department of Medical
Biochemistry, University of Oulu, Kajaanintie 52 A Fin-90220, Oulu,
Finland
FEATURES
source Location/Qualifiers
1..1615
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/sex="male"
/tissue_type="liver"
/clone_lib="Mouse adult liver cDNA library by M. Rehn"
/dev_stage="adult, 9-10 weeks"
c1..1449
/standard_name="COL18A1"

CDS

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 09:08:09 ; Search time 369.576 Seconds

(without alignments)
9529.183 Million cell updates/sec

Title: US-09-938-391-1
Perfect score: 829
Sequence: 1 cccctggcggcagatgacat.....aataaaagggaagccaaagag 829

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*1: Geneseqn1980s:*2: Geneseqn1990s:*3: Geneseqn2000s:*4: Geneseqn2001as:*5: Geneseqn2001bs:*6: Geneseqn2002s:*7: Geneseqn2003as:*8: Geneseqn2003bs:*9: Geneseqn2003cs:*10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	829	100.0	829	6	AAL46062 Canine pr
2	555	66.9	555	6	AAL46063 Canine en
3	550.4	66.4	552	3	AAL46063 Canine en
4	503.2	60.7	4875	6	ABQ54955 Human ova
5	503.2	60.7	5408	2	AAX77720 Human col
6	496.8	59.9	3394	2	AAT84484 Human alp
7	496.8	59.9	3394	2	AAX78379 Human alp
8	496.8	59.9	3394	6	ABN95680 Gene #217
9	493.2	59.5	900	6	ABK09977 Synthetic
10	493.2	59.5	4551	6	ABN85301 Human col
11	493.2	59.5	4551	6	ABV94763 Human pan
12	493.2	59.5	4551	6	ABV94763 Human pan
13	423.6	51.3	1968	3	AAC62025 Nucleotid
14	423.6	51.3	1968	6	ABQ76740 DNA encoo
15	423.6	51.3	4031	2	AAT84485 Mouse alp
16	420	50.8	816	3	AAA46013 DNA encoo
17	420	50.7	555	6	AAL66529 Human vas
18	418.4	50.6	558	6	ABA99261 Human end
19	418.4	50.5	563	2	AAX08750 Human end
20	418	50.4	546	3	AAZ27004 Human end
21	418	50.4	549	3	AAX77719 Human end
22	418	50.4	549	3	AAZ51291 Human ang
23	418	50.4	549	5	AAZ99884 Human ang
24	418	50.4	549	5	AAZ00867 Human gen

24	418	50.4	549	7	ABA00774 Human end
25	418	50.4	552	2	AAZ35375 SEQ ID 50
26	418	50.4	552	3	AAZ68203 Human end
27	418	50.4	552	3	AAZ68203 Nucleotid
28	418	50.4	552	5	AAZ88289 Human end
29	418	50.4	552	6	ABK50685 cDNA enco
30	417.2	50.3	641	4	AAH79104 Human end
31	417.2	50.3	641	6	AAH79104 Specific
32	417.2	50.3	641	6	AAH79104 Anticance
33	417.2	50.3	641	6	AAH79104 Reproduc
34	417	50.3	551	6	ABQ81193 Human end
35	412.2	49.7	6462	3	AAZ64014 Nucleotid
36	411.6	49.7	555	9	AAZ64014 Human NOV
37	410.8	49.6	534	3	AAZ27005 Alternate
38	410.8	49.6	537	5	AAZ00868 Human gen
39	398.4	48.1	650	6	ABK09978 Synthetic
40	394.2	47.6	552	3	AZ50398 Human end
41	376.8	45.5	573	3	AAZ67777 Murine en
42	376.8	45.5	573	6	AAZ47719 cDNA enco
43	376.6	45.4	565	2	AAZ58740 DNA encoo
44	376.4	45.4	580	2	AAZ08748 Mouse end
45	376.2	45.4	558	4	AAZ18701 Mouse end

ALIGNMENTS

RESULT 1
AAL46062
ID AAL46062 standard; cDNA; 829 BP.

AC AAL46062;

DT 19-JUL-2002 (first entry)

XX Canine pro-endostatin coding sequence.

XX Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
XX psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
XX corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
XX rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
XX plaque neovascularisation; telangiectasia; haemophilic joints;
XX angiofibroma; wound granulation; coronary collateral;
XX cerebral collateral; arteriovenous malformation;
XX ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
XX cytotactic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
XX gynaecological; Gene; ss.

OS Canis familiaris.

XX Key Location/Qualifiers
XX CDS 1..893
FT FT /tag= a
FT FT /product= "pro-endostatin"
FT FT /partial
FT FT /note= "no start codon"

XX EPI191036-A2.

XX 27-MAR-2002.

XX 24-AUG-2001; 2001EP-00307224.

XX 25-AUG-2000; 2000US-0227924P.

XX (PFIZ) PFIZER PROD INC.

XX Sheppard MG, Tong X; , M.

XX WPI; 2002-354068/39.

XX P-PSDB; AAO.7429.

XX An isolated nucleic acid molecule for the treatment of angiogenesis-

PT	related disorder, such as cancers or diabetic retinopathy, encodes an
XX	endostatin protein.
PS	Claim 2; Fig 2; 56pp; English.
XX	
CC	The present invention provides the protein and coding sequences of canine
CC	pro-endostatin and endostatin. The sequences can be used in the treatment
CC	and diagnosis of angiogenesis related disorders, including cancer,
CC	rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
CC	corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC	rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC	neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC	wound granulation, coronary collaterals, cerebral collaterals,
CC	arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC	neovascularisation, and fractures. The present sequence is the canine pro
CC	-endostatin coding sequence
XX	
SQ	Sequence 829 BP; 124 A; 314 C; 278 G; 113 T; 0 U; 0 Other;
	Query Match 100.0%; Score 829; DB 6; Length 829;
	Best Local Similarity 100.0%; Pred. No. 1.7e-126;
	Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 CCTGTGGGGAGATGACATCTGCGGGCCCCCGCGCTGTGTGAGACCCCGAGCCCTAC 60
DB	1 CCTGTGGGGAGATGACATCTGTGGCGGCCCGCGCTGTGTGAGACCCCGAGCCCTAC 60
QY	61 CCGCGGGCCCCGACACACGGTCTCTAGTGCATTTCCAGCGGCTCGCCCACTGTGTGG 120
DB	61 CCGCGGGCCCCGACACACGGGTCTCTAGTGCATTTCCAGCGGCTCGCCCACTGTGTGG 120
QY	121 CCGGTCCACACCCACACCCACACACAGAGCTTCAGAGTGTGTGCACCTGTGTGGCC 180
DB	121 CCGGTCCACACCCACACCCACACAGAGCTTCAGAGTGTGTGCACCTGTGTGGCC 180
QY	181 CTGAAACAGCCCGACGCGGGCGGATGCGAGGCATTCGGGGAGCGGACTTCAGTGTCTTC 240
DB	181 CTGAAACAGCCCGACGCGGGGGGATGCGAGGCATTCGGGGAGCGGACTTCAGTGTCTTC 240
QY	241 CAGCAGGCGCGCGCGGGGCTGGCGGCACCTTCGGGCGCTTCCTGTCTCGCGGCTG 300
DB	241 CAGCAGGCGCGCGCGGGGCTGGCGGCACCTTCCTGTCTCGCGGCTTCCTGTCTCGCGGCTG 300
QY	301 CAGGACCTCTTACAGATCTGTGGCGCGCGCGACCGGAGTGGCGCTGTCTGAACCTC 360
DB	301 CAGGACCTCTTACAGATCTGTGGCGCGCGCGACCGGAGTGGCGCTGTCTGAACCTC 360
QY	361 AGGGACGAGGTGCTCTTCCCGAGCTGGGAGGCTTATTCTCGGGCTCCGAGGGCCAGGTG 420
DB	361 AGGGACGAGGTGCTCTTCCCGAGCTGGGAGGCTTATTCTCGGGCTCCGAGGGCCAGGTG 420
QY	421 AAGCCGGGGCCCGCATCTTCTCTTTGACGCGAGAGATGTCCTGTGAGACACCCCGCTGG 480
DB	421 AAGCCGGGGCCCGCATCTTCTCTTTGACGCGAGAGATGTCCTGTGAGACACCCCGCTGG 480
QY	481 CCCCAGAAAGAGCGTGTGACGCGGTCCGACCCAGCGGGCGCGCGCTGACCGACAGCTAC 540
DB	481 CCCCAGAAAGAGCGTGTGACGCGGTCCGACCCAGCGGGCGCGCGCTGACCGACAGCTAC 540
QY	541 TCGGAGAGTGTGCGGACGAGAGCCCGCGCGCACCGGCGAGGCTGTGTGTGTGGCG 600
DB	541 TCGGAGAGTGTGCGGAGCGAGAGCCCGCGCGCACCGGCGAGGCTGTGTGTGTGGCG 600
QY	601 GGCAGGCTGCTGGAGAGGCGCGAGCTGCGCGCACCGCTTCGTGTGTGTCTGTGCATC 660
DB	601 GGCAGGCTGCTGGAGAGGCGCGAGCTGCGCGCACCGCTTCGTGTGTGTCTGTGCATC 660
QY	661 GAGACAGGCTCATGACCTCTTCTCCAGTGTGGGCGCGCGGCCACCGACAGCGGGGG 720
DB	661 GAGACAGGCTCATGACCTCTTCTCCAGTGTGGGCGCGCGGCCACCGACAGCGGGGG 720
QY	721 GAGGGGGCGCCGCGAGGAGCATCCGCGCGCCCCCGGGGGGCGCTGTGGCGGAGCGTTCCTG 780

Db 721 GAGGGGGCGCCGACAGGAGCATCCGCCGCCCGGGGGGGGCGCTTGCGCGGAGCGTTTGCGTG 780
 Qy 781 CACCGTCACGTTTAAATGTAATCTCTCAAGAAATAAAGAGAAAGCCAAAGAG 829
 Db 781 CACCGTCACGTTTAAATGTAATCTCTCAAGAAATAAAGAGAAAGCCAAAGAG 829
 RESULT 2
 AAL46063
 ID AAL46063 standard; cDNA; 555 BP.
 XX AAL46063;
 XX
 DT 19-JUL-2002 (first entry)
 XX
 DE Canine endostatin coding sequence.
 XX
 Kw Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
 Kw psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
 Kw corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 Kw rubecosis; Osler-Webber Syndrome; myocardial angiogenesis;
 Kw plaque neovascularisation; telangiectasia; haemophilic joints;
 Kw angiofibroma; wound granulation; coronary collateral;
 Kw cerebral collateral; arteriovenous malformation;
 Kw ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
 Kw cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
 Kw gynaecological; gene; ss.
 XX
 XX Canis familiaris.
 OS
 XX
 XX
 Ph Key Location/Qualifiers
 FT 1..555
 CDS /*tag= a
 /product= "endostatin"
 /partial
 /note= "no start codon"
 XX
 EP1191036-A2.
 XX
 XX 27-MAR-2002.
 XX
 XX 24-AUG-2001; 2001EP-00307224.
 XX
 XX 25-AUG-2000; 2000US-0227924P.
 XX
 XX (PFIZ) PFIZER PROD INC.
 XX
 XX Sheppard MG, Tong X;
 P1
 XX WPI: 2002-354068/39.
 XX P-PSDB; ARL17430.
 XX
 XX An isolated nucleic acid molecule for the treatment of angiogenesis-
 PT related disorder, such as cancers or diabetic retinopathy, encodes an
 PT endostatin protein.
 XX
 XX Claim 2; Fig 4; 56pp; English.
 XX
 XX The present invention provides the protein and coding sequences of canine
 CC pro-endostatin and endostatin. The sequences can be used in the treatment
 CC and diagnosis of angiogenesis related disorders, including cancer,
 CC rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
 CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
 CC rubecosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC wound granulation, coronary collateral, cerebral collaterals,
 CC arteriovenous malformations, ischaemic limb angiogenesis, diabetic
 CC neovascularisation, and fractures. The present sequence is the canine
 CC endostatin coding sequence
 XX
 XX Sequence 555 BP; 78 A; 203 C; 191 G; 83 T; 0 U; 0 Other;
 SQ

Query Match 66.9%; Score 555; DB 6; Length 555;

QY	541	TCGAGACGTGGCGGACCGAGAGCCCGCGGCCACCGGGCAGGCGTGGTGGCTGCTGGCG	600
Db	1906	TGTGAGACGTGGCGGACCGAGGCTCCCTCGGCCACCGGCCAGGCGCTCTGGCTGCTGGG	1965
QY	601	GGCAGGCTGCTGGAGCAGGAGGACCGGAGCTGCCGCCACGCTTCGTGGTGCTCTGCATC	660
Db	1966	GGCAGGCTCTCTGGGCGAGAGTGCOCGCGAGCTGCCCATCAGCGCTACATCGTCTGCAAT	2025
QY	661	GAGAACAGCGTATGACCTCTCTTCCAAAGTAGGCGCGCGGCCACGACAGGCGGG	720
Db	2026	GAGAACAGCTTTCATGACTCCTCCCAAGTAGCCACCGCTGGATGCAGATGCCCGGAGG	2085
QY	721	GAGGGGCGCCCGCAGGAGCATCCGCGCCCGCCGGG	756
Db	2086	ACCGCGGCTCGGAGGAAGCCCCACCGTGGGACAGGAGCAGCGCCCTGGCCC	2145
QY	757	--GGCGTGGCCGGGACGCTTGCCTGCACCG--TCACGTTTAACTGAATCCTCAGAATA	813
Db	2146	CAGGACCTGGCTGCCATCTTCTGTATAGTTCACGTTTCATGTAATCCTCAGAATA	2205
QY	814	AAAGAAAGCCAAAGAG	829
Db	2206	AAAGAAAGCCAAAGAG	2221
RESULT 7			
AAAX78379			
ID	AAAX78379 standard; cDNA; 3394 BP.		
XX	AC AAX78379;		
XX	25-AUG-1999 (first entry)		
DT	Human alpha1 (XVIII) collagen cDNA.		
DE	Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;		
KW	anti-angiogenic; heparin binding domain; receptor binding domain; mimetic;		
KW	alpha-helix A domain; carbohydrate recognition domain; CRD domain;		
KW	treatment; angiogenesis; tumour; human; ss.		
OS	Homo sapiens.		
Key	Location/Qualifiers		
FF	1..2055		
FT	/*tag= a		
FT	/product= "alpha1(XVIII) collagen"		
FT	/note= "Partial sequence, no start codon given"		
PN	W0931616-A1.		
XX	24-JUN-1999.		
XX	16-DEC-1998; 98WO-US026783.		
XX	16-DEC-1997; 97US-0069727P.		
PR	(HARD) HARVARD COLLEGE.		
XX	Olsen ER, Hohenester E, Timpl R, Sasaki T;		
PI	WPI; 1999-395243/33.		
XX	P-PSDB; AAY25113.		
DR	Identifying mimetics of mammalian endostatin.		
XX	Disclosure; Fig 5A-C; 75pp; English.		
PS	This invention describes a novel method for identifying mimetics of		
CC	mammalian endostatin. The method comprises identifying a compound having		
CC	atomic coordinates with non-trivial similarity to selected coordinates of		
CC	atoms of a mammalian endostatin involves (a) providing a library of		
CC	atomic coordinates of compounds in a library of candidate compounds, (b)		

CC comparing the library of atomic coordinates to the selected coordinates
CC of a mammalian endostatin and (c) selecting from the library at least one
CC candidate compound on the basis of selection criteria which include
CC similarities between the atomic coordinates of the selected candidate
CC compound and the atomic coordinates of the mammalian endostatin. The
CC invention also describes the use of an anti-angiogenic fragment of
CC endostatin comprising a domain selected from a heparin binding domain, a
CC receptor binding domain, and exposed on alpha-helix A domain, and a
CC carbohydrate recognition domain (CRD) domain. The methods can be used for
CC designing and selecting endostatin mimics. The compounds identified can
CC be used for treating undesired angiogenesis, e.g. tumours. This sequence
CC encodes human alpha (XVIII) collagen which is used in the description of
CC the method
XX
SQ Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 U; 0 Other;
Query Match 59.9%; Score 496.8; DB 2; Length 3394;
Best Local Similarity 77.1%; Pred. No. 2.1e-72;
Matches 660; Conservative 0; Mismatches 157; Indels 39; Gaps 3;
QY 1 CCTGCGGGGAGATGACATCTCTGCGCGCGCCCGCGCTGCGAGCCCGCCCTAC 60
DB 1378 CCTGCGGGGAGATGACATCTCTGCGCGCGCCCGCGCTGCGAGCCCGCCCTAC 1437
QY 61 CCGCGGGCCCGGACACCGGCTCTACGTGACCTTCACGCGGCTGCGCCCGCTGGTGG 120
DB 1438 CCGGAGGGCCCGACACAGCTCTACGTGACCTTCACGCGGCTGCGCCCGCTGGTGG 1493
QY 121 CCGGTCCACACCCACACACACACACACAGGACTTCCAGCTGCTGCTCAGCTGGTGGC 180
DB 1494 -----CCACCCCGCCACAGCCACCGGACTTCCAGCGGTGCTCAGCTGGTGG 1545
QY 181 CTGAACAGCCCGCAGCCGGCGGATGAGAGCATTCGCGGGAGAGGACTTCAGTGTTC 240
DB 1546 CTCAACAGCCCGCTGTGAGCGGATGCGGGGATCCCGCGGCGGAGCTTCAGTGTTC 1605
QY 241 CAGCAGCGCGCGCGCGGGGTGCGCGGACCTTCGCGGCTTCCTGCTGCGCGGTG 300
DB 1606 CAGCAGCGCGCGCGCGGGGTGCGCGGACCTTCGCGGCTTCCTGCTGCGCGGTG 1665
QY 301 CAGGAGCTCTACAGCATGCTGCGCGCGCGGACCGGAGCTGCGGTGCTCAAGCTC 360
DB 1666 CAGGAGCTCTACAGCATGCTGCGCGCGCGGAGCGGAGCTGCGGTGCTCAAGCTC 1725
QY 361 AGGAGCAGGTGCTCTTCCAGCTGCGGAGGCTTATCTCGGCTTCGAGGGCCAGCTG 420
DB 1726 AAGGAGCAGGTGCTGTTCCAGCTGCGGAGGCTTATCTCGGCTTCGAGGGTCCGCTG 1785
QY 421 AAGCCCGGGCGCGCATCTTCTTTCGACGCGAGAGATGCTCAGCAGCCCGCGCTGG 480
DB 1786 AAGCCCGGGCGCGCATCTTCTTTCGACGCGAGGAGCTCCTGAGGACCCCGACCTGG 1845
QY 481 CCGCGAAGAGCTGTGGCAGCGGCTCCGACCCCGAGGGCGCGCTCAGCAGAGCTAC 540
DB 1846 CCGCAGAGAGCTGTGGCAGCTGCTCGGACCCCGAGCGCGGAGCTGAGGAGCTAC 1905
QY 541 TCGGAGAGCTGGCGAGCGGAGCGCGCGGCGCCACCGGCGAGGCTGCTGCTGCGG 600
DB 1906 TGTGAGAGCTGGCGAGCGGAGGCTCCTCGGCGCAGGCGCGGCTCTGCTGCTGGG 1965
QY 601 GGCAGGCTGTGGAGCAGGAGCGCGAGCTGCGCGGCTTCGCTGCTGCTGCTGCTG 660
DB 1966 GGCAGGCTGTGGGCGAGAGTGGCGGAGCTGCGGCTGCGGCTGCTGCTGCTGCTG 2025
QY 661 GAGACAGCTGTGAGCTCTTCTCCAGTAGGCGCGCGCGGCGGCGGAGCGGCGG 720
DB 2026 GAGACAGCTGTGAGCTCTTCTCCAGTAGGCGCGCGCGGCGGCGGAGCGGCGG 2085
QY 721 GAGGGGCGCGCGCAGGAGCATCCGCGCGCGCGGGG----- 756
DB 2086 ACCGGCGCTCGGAGGAGCGCGGCGGCGGAGGAGCGCGGCGGCGGCGGCGGCGG 2145
QY 757 --GGGCGCTGGCGGAGCGCTTGCCTGCGCGG--TCAGGTTTATGTATCTCAAGAAATA 813

DB 2146 CAGGACCTGGTGGCTATCTTCTCTGTATAGTTCACGTTTCATGTATCTCAAGAAATA 2205
QY 814 AAAGGAAGCCAAAGAG 829
DB 2206 AAAGGAAGCCAAAGAG 2221
RESULT 8
ABN95680
ID ABN95680 standard; DNA; 3394 BP.
AC ABN95680;
XX
DT 13-AUG-2002 (first entry)
XX
XX Gene #2178 used to diagnose liver cancer.
DE Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX Homo sapiens.
XX WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1; SEQ ID NO 2178; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 U; 0 Other;
Query Match 59.9%; Score 496.8; DB 6; Length 3394;
Best Local Similarity 77.1%; Pred. No. 2.1e-72;
Matches 660; Conservative 0; Mismatches 157; Indels 39; Gaps 3;
QY 1 CCTGCGGGGAGATGACATCTCTGCGCGCGCCCGCGCTGCGAGCCCGCCCTAC 60
DB 1378 CCTGCGGGGAGATGACATCTCTGCGCGCGCCCGCGCTGCGAGCCCGCCCTAC 1437
QY 61 CCGCGGGCCCGGACACCGGCTCTACGTGACCTTCAGGCGGCTGCGCCCGCTGGTGG 120
DB 1438 CCGGAGGGCCCGACACAGCTCTACGTGACCTTCAGGCGGCTGCGCCCGCTGGTGG 1493

CC actually applied in vivo for the maturation of FPD. The use of a single
CC regulatory unit eliminates the necessity of manipulating the
CC complementary DNA coding for the FPD to isolate the segment coding for
CC the FPD, and adapt it to the expression vector. The present nucleic acid
CC sequence encodes the plasmid pBR-HR#1 FPD fusion protein of the
CC invention. The FPD fusion protein contains the mouse Ig signal peptide
CC (mIGSP) sequence fused to exons 38-41 of the human COL18A1 sequence
XX
XX
SQ Sequence 900 BP; 146 A; 336 C; 274 G; 144 T; 0 U; 0 Other;

Query Match 59.5%; Score 493.2; DB 6; Length 900;
Best Local Similarity 83.9%; Pred. No. 8.9e-72;
Matches 574; Conservative 0; Mismatches 98; Indels 12; Gaps 1;

QY 1 CCTGTGGGGGACATGACATCTCTGGGCGCGCCCGCGGCTGTGACCCCGAGCCCTTAC 60
Db 211 CCTGTGGGGGACATGACATCTCTGGGCGCGCCCGCGGCTGTGACCCCGAGCCCTTAC 270
QY 61 CCGGGGGCCCCGACACAGGCTCTTACGTGCACTTCCAGCGGCTGCGCCGACTGTGGG 120
Db 271 CCGGGAGCCCCGACACAGGCTCTTACGTGCACTTCCAGCGGCTGCGCCGACTGTGGG 326
QY 121 CCGGTCCACACCCAC 180
Db 327 -----CCACCCCGCCACACACACACACACACACACACACACACACACACACAC 378
QY 181 CTGAACAGCCCGCAGCGCGGCATCCGAGGACATCCGCGGAGCGGACTTCCAGTGTTC 240
Db 379 CTCAACAGCCCGCAGCGCGGCATCCGCGGAGCGGACTTCCAGTGTTC 438
QY 241 CAGCAGCG 300
Db 439 CAGCAGCG 498
QY 301 CAGACCTCTACAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 499 CAGACCTGTACAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 558
QY 361 AGGAGCAGGTGTCTTCCAGCTGGAGGCGCTTATTCGCGGCTTCGAGGCGCAGCTG 420
Db 559 AAGCAGCAGGTGTCTTCCAGCTGGAGGCGCTTATTCGCGGCTTCGAGGCTTCGCGT 618
QY 421 AAGCG 480
Db 619 AAGCG 678
QY 481 CCGCGAAGAGCGGTGTGCGACGCGCTCCGACCCCGAGCGCGCGCGCGCGCGCGCG 540
Db 679 CCGCAGAGAGCGGTGTGCGATGCTCGACCCCGACCGCGCGCGCGCGCGCGCG 738
QY 541 TGCAGACGTGGCGGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 739 TGTGAGACGTGGCGGACGAGGCGCTCCCTCGGCGACCGGCGCGCGCGCGCGCG 798
QY 601 GGCAGGCTGTGGAGCAGGAGGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 799 GGCAGGCTGTGGGCGAGAGTGGCGGAGCTGCGCATCAGCGCTTACATCTGTCTG 858
QY 661 GAGAACAGCGCTCATGACCTCTTC 684
Db 859 GAGAACAGCTTATGACTGCGCTC 882

RESULT 10
ABN85301
ID ABN85301 standard; DNA; 4551 BP.
XX
XX ABN85301;
XX AC
XX
XX
XX 30-SEP-2002 (first entry)
XX
XX Human collagen XVIII coding sequence.

KW Human; antirheumatic; antiarthritic; gene therapy; anti-angiogenic;
KW rheumatoid arthritis; collagen; endostatin; gene; ds.
XX Homo sapiens.
XX
XX Location/Qualifiers
FH 1..4551
FT CDS /tag= a
FT /product= "Human collagen XVIII"
XX
XX WO200253191-A1.
XX
XX 11-JUL-2002.
XX
XX 03-JAN-2002; 2002WO-KR000001.
XX
XX 05-JAN-2001; 2001KR-00000691.
XX
XX (VIRO-) VIROMED LTD.
XX
XX Kim J, Ho S, Park E, Kim S;
XX
XX WPI; 2002-583596/62.
XX P-PSDB; ABB83471.
XX
XX Novel composition for gene therapy against rheumatoid arthritis,
XX comprising a DNA encoding anti-angiogenic protein or its parts.
XX
XX Disclosure; Page 60-70; 84pp; English.

CC The present invention relates to a composition for gene therapy,
CC comprising a DNA encoding an anti-angiogenic protein, which shows
CC therapeutic effects on rheumatoid arthritis. The composition is useful
CC for treating rheumatoid arthritis and the gene therapy is effective,
CC lasting for 14 days. The present sequence is the coding sequence for
CC human collagen XVIII. Endostatin, which consists of the C-terminal 183
CC residues of collagen XVIII, was used as an anti-angiogenic protein
XX
XX Sequence 4551 BP; 780 A; 1597 C; 1522 G; 652 T; 0 U; 0 Other;

Query Match 59.5%; Score 493.2; DB 6; Length 4551;
Best Local Similarity 83.9%; Pred. No. 7.8e-72;
Matches 574; Conservative 0; Mismatches 98; Indels 12; Gaps 1;
QY 1 CCTGTGGGGGACATGACATCTCTGGGCGCGCCCGCGGCTGTGACCCCGAGCCCTTAC 60
Db 3874 CCTGTGGGGGACATGACATCTCTGGGCGCGCCCGCGGCTGTGACCCCGAGCCCTTAC 3933
QY 61 CCGGGGGCCCCGACCA CCGGCTCTTACGTGCACTTCCAGCGGCTGCGCCGACTGTGGG 120
Db 3934 CCGGGAGCCCGCACACAGCTCTTACGTGCACTTCCAGCGGCTGCGCCGACTGTGGG 3989
QY 121 CCGGTCCACACCCACACCCACACACACAGGACTTCCAGCTGCTGCTGCACTGTGGCC 180
Db 3990 -----CCACCCCGCCACACAGCCACCCGACTTCCAGCGGCTGCTGCACTGTGGG 4041
QY 181 CTGAACAGCCCGCAGCGCGGCGGCGCATGCGAGGACATCCGGGAGCGGACTTCCAGTGTTC 240
Db 4042 CTCAACAGCCCGCTGTGACGCGGCGCATGCGGCGCATCCGCGGCGGCGACTTCCAGTGTTC 4101
QY 241 CAGCAGCGCGCGCGCGGCGGCTGCGCGGCGCATCTTCCGCGGCGCTTCTGCTGCGGCTG 300
Db 4102 CAGCAGCGCGCGCGCGGCGGCTGCGGCGCATCTTCCGCGGCGCTTCTGCTGCGGCTG 4161
QY 301 CAGGACCTCTACAGCATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 4162 CAGGACCTGTACAGCATCTGTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCTC 4221
QY 361 AGGAGCAGGTGTCTTCCCGAGCTGGAGGCGCTTATTCGCGGCTTCGAGGCGGCGCTG 420
Db 4222 AAGGACGAGCTGTCTTCCAGCTGGAGGCTCTTCTTCTCAGGCTCTCAGGCTTCGCTG 4281
QY 421 AAGCGCGGCGCGCGCGCATCTTCTTCTTCGAGCGGAGAGATGTCTCTGCGAGCACCCCGCGCTG 480

Db 4282 AAGCCCGGGGACGATCTCTCTTTGACGCAAGACGCTCTGAGGACCCACCTGG 4341
 Qy 481 CCCCGAAGAGCGTGTGGACCGCTCCGACCCCGGCGCGCGCTGACCGACAGCTAC 540
 Db 4342 CCCCGAAGAGCGTGTGGACCGCTCCGACCCCGGCGCGCGCTGACCGACAGCTAC 4401
 Qy 541 TGGGAGACGCTGGCGGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
 Db 4402 TGTGAGACGCTGGCGGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4461
 Qy 601 GCGAGGCTGTGGACGAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 Db 4462 GCGAGGCTGTGGGCGAGAGTGGCGGAGCTCCATCAGCGCTACATGCTCTGCATT 4521
 Qy 661 GAGAACAGCGTATGACCTCTTC 684
 Db 4522 GAGAACAGCTTATGACTGCTTCC 4545

RESULT 11
 ID ABV94763
 XX ABV94763 standard; cDNA; 4551 BP.
 AC ABV94763;
 DT 14-JAN-2003 (first entry)
 DE Human pancreatic cancer expressed cDNA SEQ ID NO 144.
 KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
 KW cytostatic; tumour; gene; ss.
 OS Homo sapiens.
 FN WO200260317-A2.
 PD 08-AUG-2002.
 PF 30-JAN-2002; 2002WO-US002781.
 PR 30-JAN-2001; 2001US-0265305P.
 PR 31-JAN-2001; 2001US-0265882P.
 PR 09-FEB-2001; 2001US-0267568P.
 PR 21-MAR-2001; 2001US-0278651P.
 PR 28-APR-2001; 2001US-0287112P.
 PR 16-MAY-2001; 2001US-0291631P.
 PR 12-JUL-2001; 2001US-0305484P.
 PR 20-AUG-2001; 2001US-0313999P.
 PR 27-NOV-2001; 2001US-0333626P.
 XX (CORI-) CORIXA CORP.
 XX Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;
 PI WPI; 2002-627435/67.
 DR P-PSDB; ABP68617.
 XX

PT New isolated polynucleotide and pancreatic tumor polypeptides, useful for
 PT diagnosing, preventing and/or treating cancer, particularly pancreatic
 PT cancer.
 PT
 PS Claim 1; SEQ ID NO 144; 300pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (T) comprising: (a)
 CC any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)
 CC complements of (a); (c) sequences consisting of at least 20 contiguous
 CC residues of (a); (d) sequences that hybridize to (a), under moderately
 CC stringent conditions; (e) sequences having at least 75% or 90% identity
 CC to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-
 CC ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer
 CC in a patient and compositions comprising polypeptides, polynucleotides,
 CC antibodies, fusion proteins, T cell populations and antigen presenting

CC cells expressing the polypeptide are useful in treating pancreatic cancer
 CC and stimulating an immune response. The polynucleotides can be used as
 CC probes or primers for nucleic acid hybridisation, in the design and
 CC preparation of ribozyme molecules for inhibiting expression of the tumour
 CC polypeptides and proteins in the tumour cells, in vaccines and for gene
 CC therapy. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

SEQ Sequence 4551 BP; 780 A; 1597 C; 1522 G; 652 T; 0 U; 0 Other;

Query Match 59.5%; Score 493.2; DB 6; Length 4551;
 Best Local Similarity 83.9%; Pred. No. 7.8e-72;
 Matches 574; Conservative 0; Mismatches 98; Indels 12; Gaps 1;

Qy 1 CCTTGGCGGCGACATGACATCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 60
 Db 3874 CCTTGGCGGCGACATGACATCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 3933
 Qy 61 CCGGGGCGTGG 120
 Db 3934 CCGGGGCGTGG 3989
 Qy 121 CCGGTCCACACCCACACCCACACCCACACCCACACCCACACCCACACCCACACCC 180
 Db 3990 -----CCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4041
 Qy 181 CTGAACAGCGTTC 240
 Db 4042 CTGAACAGCGTTC 4101
 Qy 241 CAGCAGCGTGG 300
 Db 4102 CAGCAGCGTGG 4161
 Qy 301 CAGGACCTCTACAGCATCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGTCAACCTC 360
 Db 4162 CAGGACCTCTACAGCATCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGTCAACCTC 4221
 Qy 361 AGGACGAGGCTCTCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGG 420
 Db 4222 AGGACGAGGCTCTCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGG 4281
 Qy 421 AAGCCCGGCGTGG 480
 Db 4282 AAGCCCGGCGTGG 4341
 Qy 481 CCCCGGAAGAGCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGG 540
 Db 4342 CCCCGGAAGAGCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGG 5401
 Qy 541 TGGGAGACGCTGGCGGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGCG 600
 Db 4402 TGTGAGACGCTGGCGGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGCG 4461
 Qy 601 GCGAGGCTGTGGGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGCG 660
 Db 4462 GCGAGGCTGTGGGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGCG 4521
 Qy 661 GAGAACAGCGTATGACCTCTTC 684
 Db 4522 GAGAACAGCTTATGACTGCTTCC 4545

RESULT 12
 AAC62025
 ID AAC62025 standard; DNA; 968 BP.
 XX
 AC AAC62025;
 XX
 DT 06-MAR-2001 (first entry)
 XX
 DE Nucleotide sequence of vaa-endostatin fusion protein in pANT3052.

XX Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
KW endostatin; cancer; tumour growth; angiogenesis; ss.
XX
OS Synthetic.
OS Streptomyces sp.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_binding 220..234
FT /tag= a
FT /note= "proposed SnpR binding site"
FT RBS 307..311
FT /tag= b
FT CDS 318..968
FT /tag= c
FT sig_peptide 318..401
FT /tag= d
FT /note= "vaa signal sequence"
FT mat_peptide 402..968
FT /tag= e
FT /note= "endostatin"
PN WO20060945-A1.
XX
PD 19-OCT-2000.
XX
PF 12-APR-2000; 2000WO-US009747.
XX
PR 13-APR-1999; 99US-0129084P.
XX
PA (MERI) MERCK & CO INC.
XX
PI Desanti CL, Strohl WR;
XX
XX WPI; 2000-686970/67.
DR P-PSDB; AAB30495.
XX
XX Preparation of soluble recombinant endostatin involves transforming
PT Streptomyces host with expression vector comprising nucleotide sequence
PT encoding endostatin operably linked to linker and leader peptide.
XX
PS Example 1; Fig 10A-B; 57pp; English.
XX
XX The present sequence encodes a fusion protein of vaa and endostatin. The
CC specification describes a method for the production of soluble,
CC recombinant human endostatin in Streptomyces. Leader sequences of
CC Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins
CC are linked to the N-terminal of endostatin. This ensures that endostatin
CC protein is produced as a secreted, soluble protein which needs no
CC refolding, is stable in the fermentation broth and is produced in large
CC quantities. The method is used for preparing soluble recombinant human,
CC murine or primate endostatin, which is useful in the treatment of cancer,
CC inhibition of tumour growth, inhibition of angiogenesis, isolation of
CC receptors for endostatin and for identification of anti-angiogenic
CC compounds in assays. The endostatin protein is produced as a secreted,
CC soluble protein which needs no refolding, is stable in the fermentation
CC broth and is produced in large quantities. Streptomyces are amenable
CC for cultivation in large fermentations allowing for large quantities of
CC soluble endostatin to be produced
XX
SQ Sequence 968 BP; 153 A; 335 C; 331 G; 149 T; 0 U; 0 Other;

Query Match 51.3%; Score 425.6; DB 3; Length 968;
Best Local Similarity 83.0%; Pred. No. 9.1e-61;
Matches 485; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 101 CGGCTGCGCCCACTGTGGCGCGTCCACACCCACACCCACACGAGACTTCCAGC 160
Db 379 CGGCAACGCGCGCGGCGAGCGCGTCCCGCGGATCCACAGCCACCGCGACTTCCAGC 438
Qy 161 TGGTGTGCACTGTGGCGTGAACAGCCCGCAGCGCGCGCATGGAGCATCCCGGG 220

Db 439 CGGTGCTCCACTGGTTGGCTCAACAGCCCCCTGTGTCAGGGGCATGCGGGCATCGCG 498
Qy 221 GAGCGACTTCCAGTGTCTTCAGCAGCGCGCGCGCGGGCTGGCCGGCACTTCGCGG 280
Db 499 GGGCGGACTTCCAGTGTCTTCAGCAGCGCGCGCGCGGGCTGGCCGGCACTTCGCGG 558
Qy 281 CTTTCTGCTGCTGCGGCTGCGAGACCTCTACAGCATCGTGCAGCGCGCGCGCACCG 340
Db 559 CTTTCTGCTGCTGCGGCTGCGAGACCTGTACAGCATCGTGCAGCGCGCGCGCACG 618
Qy 341 GGGTCCCGCTGCTCAACCTCAGGACGAGTGTCTTCCCGAGCTGGAGGCTTATCT 400
Db 619 CCGTGCCCATGCTCAACCTCAAGCAGAGTGTCTTCCAGCTGGAGGCTCTGTCT 678
Qy 401 CGGCTCCAGGGCCAGCTGAAGCCCGGGCCCGCATCTCTTCTTCAGCGGAGATG 460
Db 679 CAGGCTCTGAGGGTCCGCTGAAGCCCGGGCCAGCATCTTCTCTTTGACGGGAGACG 738
Qy 461 TCCTGCAGCAGCCCGCTGCGCCCGGAGAGCGTGTGGCAGCGTCCGACCCAGCGGGC 520
Db 739 TCCTGAGGACACCCACCTGCGCCCGAGAGAGCGTGTGGCATGGCTCGGACCCCAACGGC 798
Qy 521 GCGGCTGACCGACAGCTACTGCGAGAGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 580
Db 799 GCAGGCTGACCGAGAGCTACTGTCGAGAGCTGGCGGAGGAGGAGGAGGAGGAGGAGG 858
Qy 581 AGGCGTCTGCTGCTGCGCGGCGAGGCTGTGGAGAGGAGGAGGAGGAGGAGGAGGAGG 640
Db 859 AGGCTCTCTGCTGCTGGGGGCGAGGCTCTTGGGGGAGAGTGGCGGAGTGCATCAG 918
Qy 641 CCTTCTGCTGCTGCTGCTGATCGAGAACAGCGTCACTGACCTCTTC 684
Db 919 CCTACATGCTGCTGCTGCTGATGAGAACAGCTTCACTGACTGCTCC 962

RESULT 13
ABQ76740
ID ABQ76740 standard; DNA; 1564 BP.
AC ABQ76740;
XX
DT 03-MAR-2003 (first entry)
XX
XX DNA encoding human Endostatin/IgG1Fc fusion construct.
XX Human; endostatin; IgG1Fc; tumour; vascular endothelial proliferation;
KW vascular endothelial cytopoiesis inhibiting factor; inhibitor;
KW fusion construct; ds.
XX Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FT CDS 1..1527
FT /tag= a
FT /product= "endostatin/IgG1Fc construct"
FT /transl_except= (pos58..60,aa:Ala)
FT /transl_except= (pos:811..816,aa:GGGSGG)
XX
XX CM1354186-A.
XX
PD 19-JUN-2002.
XX
PF 30-NOV-2000; 2000CN-00123347.
XX
PR 30-NOV-2000; 2000CN-00123347.
XX
XX (LIAO-) LIAONING WEIXING BIOLOGICAL PROD INST CO.
XX Chen L, Li Z, Liu Q;
XX WPI; 2002-751441/82.
DR P-PSDB; ABG73586.
DR

XX Preparation of recombinant human vascular endothelial cytopoiesis
 PT suppressor factor with human IgG1Fc fragment molecular structure and
 PT application of its product.
 XX
 XX Disclosure; Page 7-8 (Disclosure); 12pp; Chinese.
 XX
 XX This invention describes a novel method for the preparation of
 CC recombinant human vascular endothelial cytopoiesis inhibiting factor with
 CC human IgG1Fc fragment molecular structure and its product application.
 CC The novel factor is derived from endostatin (using PCR to screen a human
 CC foetal kidney cell cDNA library) and human IgG1Fc. The product of the
 CC invention can specifically inhibit tumour vascular endothelial
 CC proliferation and can be used for curing several tumour types. This
 CC sequence encodes a fusion construct composed of human endostatin and
 CC human IgG1Fc, described in the disclosure of the invention
 XX
 XX Sequence 1564 BP; 345 A; 491 C; 436 G; 291 T; 0 U; 0 Other;

Query Match 51.1%; Score 423.6; DB 6; Length 1564;
 Best Local Similarity 77.9%; Pred. No. 1.9e-60;
 Matches 528; Conservative 0; Mismatches 139; Indels 11; Gaps 1;

QY 137 CCACACCCACAGGACTCCAGCTGGTGGTGGACCTGGTGGCCCTGAACAGCCGCGCAGC 196
 DB 260 CTCACAGCCACCGGACTTCAGCCGGTGTCTCCACTGGTGGCTCAAGCCCCCTGT 319
 QY 197 CGGCGGCTATGCGAGGCTATCCGGGAGCGGACTTCCAGTCTTCCAGAGCGCGCGCG 256
 DB 320 CAGGCGGCTATGCGGCGATCCCGGGCGGACTTCCAGTCTTCCAGAGCGCGCGCG 379
 QY 257 CGGGGCTGCGCGGACCTTCGGGGCTTCCGCTGCTGGCGCTGCGAGGCTTCCAGCA 316
 DB 380 TGGGGCTGCGGGGACCTTCGGGGCTTCCGCTGCTGGCGCTGCGAGGCTTCCAGCA 439
 QY 317 TGTGCGCGCGCGGACCGGACCGGGGTCGCGCTGCTCAACTCAGGAGCGAGTCTCT 376
 DB 440 TGTGCGCGCGTCCGACCGCGGCGGCTGCGCTCAACTCAGGAGCGAGTCTCT 499
 QY 377 TCCGAGCTGGGAGGCTTATCTCGGCTCCGAGGCGGAGCTGAAGCGCGCGCGCA 436
 DB 500 TTCCAGCTGGGAGGCTTGTCTCAGGCTTGAGGCTCCGCTGAAGCGCGGCGGCA 559
 QY 437 TCTTCTTTTCCAGCGGAGATGCTCTGAGCAACCCGCTGCGCGCGGAGAGCGTGT 496
 DB 560 TCTTCTTTTCCAGCGGAGGAGCTCTGAGGCAACCCGCTGCGCGCGGAGAGCGTGT 619
 QY 497 GCGAGGCTCCGACCCGCGGCGCGCTGAGCGGAGCTTACTGCGAGAGCGTGGCGA 556
 DB 620 GCGATGGCTCGGACCCGCGGCGGAGCTGAGCGGAGCTTACTGAGAGCGTGGCGA 679
 QY 557 CGGAGGCGCGCGCGGCGGCGGCGGAGCGGCTGCTGCTGCTGCGGCGGAGCTGCGAGC 616
 DB 680 CGGAGGCTCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 739
 QY 617 AGGAGGCGCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 676
 DB 740 AGAGTGGCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 799
 QY 677 CTTCTTTTCCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 736
 DB 800 CTGCGCTCCAGTAGTAGG-----AGGAAGCGGAGGTGGAGGCTCCGTCGACAA 848
 QY 737 GAGCATCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 796
 DB 849 AACTCAGACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 908
 QY 797 GTAATCTTCAAGAAATAA 814
 DB 909 CTTCCCGCGCAAAACCAA 926

RESULT 14

AAT84485
 ID AAT84485 standard; cDNA; 4031 BP.
 AC AAT84485;
 XX
 XX 19-NOV-1997 (first entry)
 XX
 XX Mouse alpha-1 collagen cDNA.
 XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration; ss.
 XX Mus musculus.
 XX
 XX Key Location/Qualifiers
 CDS 1..3867
 /tag= a
 XX
 XX US5643783-A.
 XX
 XX 01-JUL-1997.
 XX
 XX 01-DEC-1993; 93US-00159784.
 XX
 XX 01-DEC-1993; 93US-00159784.
 XX (HARD) HARVARD COLLEGE.
 XX
 XX Oleen BR, Oh SP;
 XX
 XX WPI; 1997-350247/32.
 XX P-PSDB; AAW26328.
 XX
 XX Nucleic acid encoding human alpha-1 collagen - for production of
 PT recombinant alpha-1 collagen, for use in the treatment of cartilage
 PT degeneration.
 XX
 XX Disclosure; Fig 2; 35pp; English.
 XX
 XX A cDNA clone (AAT84485) codes for a mouse novel type alpha-1 (XVIII)
 CC collagen (AAW26328) that is expressed in multiple organs, especially the
 CC liver, lung and kidney. It was isolated by screening murine 15.5- and
 CC 17.5-day embryo libraries with probes based on murine type XII collagen
 CC and on human alpha-1 collagen cDNA. Isolated nucleic acids can be used to
 CC express recombinant alpha-1 collagen in transformed host cells. Claimed
 CC nucleic acid (see AAT84484) coding for human alpha-1 collagen (AAW26327)
 CC can be used to treat patients suffering from diseases associated with
 CC degradation of cartilage, and for supplementing collagen
 XX
 XX Sequence 4031 BP; 873 A; 1172 C; 1266 G; 720 T; 0 U; 0 Other;
 Query Match 51.0%; Score 423; DB 2; Length 4031;
 Best Local Similarity 76.9%; Pred. No. 2.2e-60;
 Matches 548; Conservative 0; Mismatches 150; Indels 15; Gaps 2;
 QY 1 CCTGCGCGGAGATGACATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAC 60
 DB 3184 CCTGCGGAGGAGATGACATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTAC 3243
 QY 61 CCGCGGG---CCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCT 117
 DB 3244 CCGGAGGTTCCACATCACACAGTTCCTATGTGACCTGCGCGGCGGCGGCGGCGGCGG 3299
 QY 118 GGGCGCGTCCACACCCAC 177
 DB 3300 -----CCTCTACTGTCTCATCTCATCAGGACCTTCAGGAGTGTCTCAGCTTGGTG 3351
 QY 178 GCGCTGAACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 237
 DB 3352 GCACTGAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3411
 QY 238 TTCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 297
 DB 3412 TTCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3471

61 CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 120
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1438
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 121
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1494
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 181
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1545
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 240
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1546
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 241
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1546
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 300
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1565
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 301
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 360
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 420
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1725
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 420
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1785
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 421
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 480
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1786
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 540
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1845
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 600
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1905
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 660
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 2025
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 720
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 2085
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 756
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 2145
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 813
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 2205
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 829
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566
Qy CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 2221
Db CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 1566

RESULT 2

US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4031
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-159-784-1

Query Match 51.0%; Score 423; DB 1; Length 4031;
Best Local Similarity 76.9%; Pred. No. 2.3e-72;
Matches 548; Conservative 0; Mismatches 150; Indels 15; Gaps 2;
Qy 1 CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 60
Db 3184 CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3243
Qy 61 CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 117
Db 3244 CCGGGGCGCCGACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3299
Qy 118 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 177
Db 3300 CCGGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3351
Qy 178 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 237
Db 3352 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3411
Qy 238 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 297
Db 3412 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3471
Qy 298 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 357
Db 3472 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3531
Qy 358 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 417
Db 3532 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3591
Qy 418 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 477
Db 3592 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3651
Qy 478 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 537
Db 3652 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3711
Qy 538 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 597
Db 3712 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 3771
Qy 598 GGGGCGCCGTCACACACGCTCTACGCTCCAGCGGCTCGCCCACTGCTGG 657

; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
; FILE REFERENCE: 05213-0370
; CURRENT APPLICATION NUMBER: US/09/206,059
; CURRENT FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-206-059-30

Query Match 50.4%; Score 418; DB 3; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.8e-71;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 139 CACACCCACAGACTTCAGCTGTGTGCTGACCTGGTGGCCCTGAACAGCCCGCAGCG 198
Db 1 CACAGCCACCGGACTTCAGCGGGTGTCCACCTGGTGGCTCAACAGCCCGCTGTCA 60
QY 199 GGGCGCATCGAGGACATCCGGGAGCGGACTTCCAGTGTCTCCAGAGGGCGCGCGCG 258
Db 61 GCGCGCATCGGGGATCCGCGGGGCGGACTTCCAGTGTCTCCAGAGGGCGCGCGCGTG 120
QY 259 GGCTGGCGGACATTCGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318
Db 121 GGGCTGGCGGACATTCGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 319 GTGCGCGCGCGGACCGGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 378
Db 181 GTGCGCGCGCGGACCGGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 379 CCAGCTGGGAGGCTTATTCGCGGCTCCGAGGCGGAGCTGAGAGCGGCGGCGGCGGCGG 438
Db 241 CCAGCTGGGAGGCTTATTCGCGGCTCCGAGGCGGAGCTGAGAGCGGCGGCGGCGGCGG 300
QY 439 TTCTCTTTTCAGCGGAGAGATGTCTTCAGGACCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 498
Db 301 TTCTCTTTTCAGCGGAGAGATGTCTTCAGGACCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 499 CAGCGTTCGAGCCCGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 558
Db 361 CAGCGTTCGAGCCCGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 559 GAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 618
Db 421 GAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 619 GAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 678
Db 481 AGTGGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
QY 679 TCCTTC 684
Db 541 GCCTCC 546

RESULT 6
US-09-315-689-6
; Sequence 6, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 534

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-315-689-6
Query Match 49.6%; Score 410.8; DB 4; Length 534;
Best Local Similarity 85.6%; Pred. No. 4.3e-70;
Matches 457; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 151 GACTTCCAGCTGGTGTGCTGACCTGGTGGCCCTGAACAGCCCGCAGCGCGGCGGATGCGA 210
Db 1 GACTTCCAGCGGTGTGCTGACCTGGTGGCCCTGAACAGCCCGCAGCGCGGCGGATGCGG 60
QY 211 GGCATCCGGGAGCGGACTTCCAGTGTCTTCCAGGAGCGCGCGCGCGGCTGCGCGGCG 270
Db 61 GGCATCCGGGAGCGGACTTCCAGTGTCTTCCAGGAGCGCGCGCGCGGCTGCGCGGCG 120
QY 271 ACCTTCCGGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 330
Db 121 ACCTTCCGGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 331 GACGCGACCGGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 390
Db 181 GACGCGACCGGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 391 GCCTTATTCGCGGCTCCGAGGCGGAGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 450
Db 241 GCCTTATTCGCGGCTCCGAGGCGGAGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY 451 GGCAGAGATGTCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 510
Db 301 GGCAGAGATGTCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 511 CCCAGCGGCGGCGGCTGCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 570
Db 361 CCCAGCGGCGGCGGCTGCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 571 GCCACCGGCGGCGGCTGCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 630
Db 421 GCCACCGGCGGCGGCTGCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 631 TGCGCGCGGCGGCTGCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 684
Db 481 TGCGCGCGGCGGCTGCTGCGAGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 534

RESULT 7
US-09-561-500-12
; Sequence 12, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip B. Thorpe
; APPLICANT: Rolf A. Brecken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-561-500-12
Query Match 45.5%; Score 376.8; DB 4; Length 573;
Best Local Similarity 79.3%; Pred. No. 1.3e-63;

[illegible]

```

RESULT 9
US-09-561-526-12
/ Sequence 12, Application US/09561526
/ Patent No. 6416758
/ GENERAL INFORMATION:
/ APPLICANT: Philip E. Thorpe
/ APPLICANT: Rolf A. Brekken
/ TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
/ FILE REFERENCE: 4001.002586
/ CURRENT APPLICATION NUMBER: US/09/561,526
/ CURRENT FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/131,432
/ PRIOR FILING DATE: 1999-04-28
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 12
/ LENGTH: 573
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
/ OTHER INFORMATION: OLIGONUCLEOTIDE
/ NAME/KEY: CDS
/ LOCATION: (1)..(573)
US-09-561-526-12

```

Matches 447; Conservative 0; Mismatches 117; Indels 0; Gaps 0

Qy 126 CCACACCACACCACCCACCCAGGACTTCAGCTGGGTGCACCTGGTGGCCCTGAA 185

Query Match	45.5%	Score 376.8	DB 4	Length 573
Best Local Similarity	79.3%	Pred. No. 1.3e-67		
Matches 447	Conservative	0	Mismatches 117	Indels 0
QY	126	CCACACCCACACCCACCCACCCAGGAGTTCACAGCTGGTCTGCACCTGCTGGCCCTGAA	185	
Db	9	CCATCACCATCACCATATCTATCAGGACTTTCAGCCAGTCTCCACTGCTGGCACTGAA	68	

Query Match	45.5%	Score 376.8	DB 4	Length 573
Best Local Similarity	79.3%	Pred. No. 1.3e-63		
Matches 447	Conservative 0	Mismatches 117	Indels 0	Gaps 0
QY	126	CCACACCCACACCCACACCCACGAGCATTCACAGCTGGTGTGCACCTGTGGCCCTGAA	185	
Db	9	CCATCACCATCACCATATCTCATCAGAGCTTTTCAGCCAGTGTCCACTGTGGCACTGAA	68	
QY	186	CAGCCCGCAGCCGGCGGCGCATCGGAGSCATCGGGGAGCGGAGCTTCAGCTGTTTCCAGCA	245	

```

Db 69 CACCCCTGCTGAGGAGGATCGGTATCCGTGAGCAGATTTCCAGTGTCTCCAGCA 128
Qy 246 GGGCGCGCGCGCGCGCGCTGCGCGCACCTTCCTCGGGCTTCCTGTCTCGCGCTGACGA 305
Db 129 AGCCGAGCGCTGGGCTGTGGGCGACCTTCCTGGGCTTCCTGTCTCTAGGCTGACGA 188
Qy 306 CTTCTACAGATCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
Db 189 TCTCTATAGCATCTGTGGCGCGCTGTGACCGGGGTCTGTGGCCATCGTCAACCTGAAGA 248
Qy 366 CGAGGTGCTCTTCCCGAGCTGGGAGGCTTATTTCTGGGCTCCGAGGGCGAGCTGAAGCC 425
Db 249 CGAGGTGCTATCTCCCGAGCTGGGAGCTCCCTGTTTCTGGCTCCGAGGTCGAAGCA 308
Qy 426 CGGGCGCGCATCTCTCTTTGACCGCGAGAGATGCTCTGAGCAGACCGCGCTGCGCGCG 485
Db 309 CGGGCGCGCATCTCTCTTTGACCGCGAGAGATGCTCTGAGCAGACCGCGCTGCGCGCG 368
Qy 486 GAAGAGGCTGTGCGAGGCTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545
Db 369 GAAGAGGCTATGCGAGCGCTGCGACCGCGAGTGGCGGAGGCTGATGAGAGTTACTGTGA 428
Qy 546 GAGTGGCGGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605
Db 429 GACATGCGGAAGTGAAGTACTGGGCTACAGGTCAAGGCTCTCTGCTGTCTGAGGCGAG 488
Qy 606 GCTGCTGAGCAGAGCGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
Db 489 GCTCTTGGAGAGAGAGTGGAGGCTGCGAGCTGCGCAGCAGTACATCTGCTGTGATGAGA 548
Qy 666 CAGGCTCATGACCTCTCTTCTCCAA 689
Db 549 TAGCTTCATGACCTCTTCTCCAA 572

```

```

RESULT 12
US-08-985-526-37
; Sequence 37, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985.526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-985-526-37

Query Match
Best Local Similarity 45.4%; Score 376.6; DB 3; Length 565;
Matches 445; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 135 CACCCACACCCACCCAGGACTTCCAGCTGGTGTGACCTGGTGGCCCTGAACAGCCGCA 194
Db 7 CATGCATATCATAGGACTTTCAGCCAGTGTCTCCACCTGGTGGACCTGAACACCCCT 66
Qy 195 GCGGGCGCGCATCGAGGCACTCCGGGAGCGGACTTCCAGTGTCTTCCAGCAGGCGCGCG 254
Db 67 GTCTGAGGACATCGGTGTATCCGTGGAGCAGATTTCCAGTGTCTTCCAGCAAGCCGAGC 126
Qy 255 CGCGGGCTGGCGGCACTTCCGGGCTTCTGTCTGCTGCGGCTGCGAGGACCTCTACAG 314
Db 127 CGTGGGCTGTGGGCACTTCCGGGCTTCTGTCTTCTAGGCTGCGAGGATCTCTATAG 186
Qy 315 CATCGTGGCGCGCGCGAGCGGCGGCTGCTGCTGAGCAGCAGCGGCTGCTCAACCTCAGGGAGCGAGTCT 374
Db 187 CATCTGCGCGCTGTGACCGGGGTCTGTGCGCATCTCAACCTGAAGGAGAGGTCT 246
Qy 375 CTTCCCGAGCTGGAGGCTTATTTCTGGGCTTCCAGGCGCAGCTGAAGCCCGGGCGCG 434
Db 247 ATCTCCAGCTGGGACTCTCTTTCTGGCTCCAGGCTCAACTGCAACCGGGGCGCG 306
Qy 435 CATCTTCTCTTTCGAGCGGAGAGATGCTCTGAGCAGCAGCGGCTGCGGCTGCGGCGGAGAGCGT 494
Db 307 CATCTTTTCTTTGAGCGGAGAGATGCTCTGAGCAGCAGCGCTGCGGCGGAGAGCGT 366
Qy 495 GTGGCAGCTCGACCCCGCGCGCGCGCGCTGCTGAGCAGCAGCGGCTGCTGAGCAGCTGCGGAG 554
Db 367 ATGGCAGCGCTCGACCCCGCGCGCGGCTGAGGAGGCTGATGAGAGGATTTCTGTGAGATGCGG 426
Qy 555 GACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGAGCAGCAGCGGCTGCGGCGGAGAGCGT 614
Db 427 AACTGAACTACTGGGCTTACAGGTCAGGCTGCGGCTTCTGCTGCTGAGAGAGTCTGAGAGATGCGG 486
Qy 615 GCAGGAGCGCGCGAGCTGCGGCGCGCGCGCTGCTGAGCAGCAGCGGCTGCTGCTGAGAGAGAGCGT 674
Db 487 ACAGAAAGCTGCGAGTGGCGGCTGCGCAGCAGCTGAGTCTGCTGCTGAGAGAGAGTCTGAG 546
Qy 675 GACCTCTTCTCCAAAGTAG 693
Db 547 GACCTCTTCTCCAAATAG 565

```

```

RESULT 13
US-09-231-077D-7
; Sequence 7, Application US/09231077D
; Patent No. 6653098
; GENERAL INFORMATION:
; APPLICANT: Harding, B.I.
; APPLICANT: Vieland, B.N.
; TITLE OF INVENTION: Method of producing mouse and human
; TITLE OF INVENTION: endostatin
; FILE REFERENCE: S03071-00-US
; CURRENT APPLICATION NUMBER: US/09/231,077D
; CURRENT FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/075,587
; PRIOR FILING DATE: 1998-02-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 580
; TYPE: DNA
; ORGANISM: human
; US-09-231-077D-7

Query Match
45.4%; Score 376.4; DB 4; Length 580;

```

RESULT 14
 US-09-449-293-3
 ; Sequence 3, Application US/09449293
 ; Patent No. 6267954
 ; GENERAL INFORMATION:
 ; APPLICANT: Abitbol, Marc
 ; APPLICANT: Uteza, Yves
 ; APPLICANT: Menasche, Maurice
 ; APPLICANT: Bossard, Carine
 ; APPLICANT: Van Den Bergh, Loic
 ; APPLICANT: Bonnel, Sebastien
 ; APPLICANT: Prats, Hervé
 ; APPLICANT: Honiger, Jiri
 ; APPLICANT: Neuner-Jehle, Martin
 ; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
 ; FILE REFERENCE: 8076.202US01
 ; CURRENT APPLICATION NUMBER: US/09/449,293
 ; CURRENT FILING DATE: 1999-11-24
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 558
 ; TYPE: DNA
 ; ORGANISM: Rattus rattus
 US-09-449-293-3

Best Local Similarity	79.7%;	Pred. No. 1.7e-63;	
Matches	444;	Conservative	0; Mismatches 113; Indels 0; Gaps 0

RESULT 15
US-09-775-325-3
; Sequence 3, Application US/09775325
; Patent No. 650049
; GENERAL INFORMATION:
; APPLICANT: Abitbol, Marc
; APPLICANT: Uteza, Yves
; APPLICANT: Menasche, Maurice
; APPLICANT: Bossard, Carine
; APPLICANT: Van Den Berghie, Loic
; APPLICANT: Bonnel, Sebastian
; APPLICANT: Prats, Hervé
; APPLICANT: Honiger, Jiri
; APPLICANT: Neuner-dehite, Martin
; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
; FILE REFERENCE: 8076.202USD1
; CURRENT APPLICATION NUMBER: US/09/775,325
; CURRENT FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: US 09/449,293
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Rattus rattus
US-09-775-325-3

Best Local Similarity 79.7%; Pred. NO. 1.7E-637;
Matches 444; Conservative 0; Mismatches 113; Indels 0; Gaps 0

Qy	137	CCACACCCACAGGACTTCCAGCTGGTGTGCTGCACTGTGTGGCCCTGAACAGCCCGCAGC	196
Db	2	CTCATACTATCAGGACTTTACGCCAGTGTCTCCACTGTGTGGCACTGAACACCCCTGT	61
Qy	197	CGGGCGGCATGCGAGGACATCCGGGAGCGGACTTCCAGTGTTCAGCAAGCGCGCGCG	256
Db	62	CTGGAGGCATGCGTGTATCCGTGGAGCAGATTTCCAGTGTTCAGCAAGCCCGAGCCG	121
Qy	257	CGGGGTGGCCGGCACCTTCCGGGCTTCTGTCTGCTGGCTGAGGACCTCTACAGCA	316
Db	122	TGGGGTGTGCGGCACCTTCCGGGCTTCTGTCTGCTGCTAGCTGAGGATCTCTATAGCA	181
Qy	317	TGCTGCGCGCGCCGACCCGACCCGGGGTCCCGTGTCTCAACTCAGGACGAGTGTCT	376
Db	182	TGCTGCGCGTGTGACCGGGGCTGTGCCCATGCTCAACCTGAAGCAGAGTGTCTAT	241
Qy	377	TCCCGAGCTGGGAGGCTTTATTCTGGGCTCCGAGGCGCAGCTGAAGCCCGGGCCCGCA	436
Db	242	CTCCAGCTGGGACTCCCTGTTTCTGGGTCCCGGGTCACTGCAACCCGGGCGCCGCA	301
Qy	437	TCCTTCTTTTCGACGGCAGAGATGCTCTGACACCCCGCCTTGGCCCGGAGAGCGTGT	496
Db	302	TCCTTTCTTTTGACGGCAGAGATGCTCTGAGACACCCAGCCTGGCCCGAGAGAGCGTAT	361
Qy	497	GGACGGCTCCGACCCCGAGGGCGCCGCTGACCGACAGCTACTCGAGACGTGGCGGA	556
Db	362	GGCACGGCTCGGACCCCGAGGGCGAGGCTGATGGAGAGTTACTGTGAGACATGGCGAA	421
Qy	557	CGGAGGCCCGGGGCGCACCGGGCAGGCGTCTGCTGCTGGCGGGCAGGCTGTGGAGC	616
Db	422	CTGAATACTTGGGGCTACAGTCAAGGCTCTCCCTGCTGTGAGGAGGCTCTTGGAC	481
Qy	617	AGGAGCGCGAGCTGCGGCCAGCCCTTCGTGGTGTCTGCTGATCGAGAACCGGTCTATGA	676
Db	482	AGAAAGCTGGAGCTGCCACACAGCTACATGCTCTGTGCTTGAATAGCTTCTATGA	541
Qy	677	CCTCCTTCTCCAGTAG	693
Db	542	CCCTTCTCCAAATAG	558

Search completed: March 29, 2004, 13:41:19
 Job time : 89.8584 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 13:38:55 ; Search time 325.85 Seconds
(without alignments)
9471.775 Million cell updates/sec

Title: US-09-938-391-1

Perfect score: 829
Sequence: 1 ccttgcggcagatgacat.....aataaaggagcaagag 829

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	829	US-09-938-391-1	Sequence 1, Appli
2	555	66.9	555	US-09-938-391-3	Sequence 3, Appli
3	550.4	66.4	552	US-10-131-241-50	Sequence 50, Appl
4	550.4	66.4	552	US-10-292-418-34	Sequence 34, Appl
5	503.2	60.7	4875	US-10-264-049-835	Sequence 835, App
6	496.8	59.9	3394	US-09-880-107-2178	Sequence 2178, Ap
7	493.2	58.5	4551	US-10-060-036-144	Sequence 144, App
8	418	50.4	546	US-10-042-347-4	Sequence 4, Appli
9	418	50.4	549	US-10-131-241-53	Sequence 53, Appl
10	418	50.4	549	US-10-292-418-3	Sequence 3, Appli
11	418	50.4	552	US-09-873-676-30	Sequence 30, Appl
12	417	50.3	551	US-10-080-797-2	Sequence 2, Appli
13	415.6	50.1	632	US-10-131-241-51	Sequence 51, Appl
14	411.6	49.7	555	US-10-210-172-161	Sequence 161, App
15	410.8	49.6	534	US-10-042-347-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-938-391-1

; Sequence 1, Application US/09938391

; Publication No. US20030158099A1

; GENERAL INFORMATION:

; APPLICANT: Tong, et al.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING

; FILE REFERENCE: PC10790A

; CURRENT APPLICATION NUMBER: US/09/938,391

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 1

; LENGTH: 829

; TYPE: DNA

; ORGANISM: CANINE PRO ENDOSTATIN NUCLEOTIDE SEQUENCE

US-09-938-391-1

Query Match 100.0%; Score 829; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 1.7e-186;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CCCTGCGGCGAGATGACATCTGCGCGGCCCCCGCGCGCTGTGTGACCCCGACCCCTAC	60
Db	1	CCCTGCGGCGAGATGACATCTGCGCGGCCCCCGCGCGCTGTGTGACCCCGACCCCTAC	60
Qy	61	CCCGGGGGCCCGACACCGGCTCTAGTGCACATTCACAGCGCGCTGCGCCACTGGTGGG	120
Db	61	CCCGGGGGCCCGACACCGGCTCTAGTGCACATTCACAGCGCGCTGCGCCACTGGTGGG	120
Qy	121	CCCGTCCACACCCACACCCACACCCACAGACTTCAGTGTGTGTGACCTGGTGGCC	180
Db	121	CCCGTCCACACCCACACCCACACCCACAGACTTCAGTGTGTGTGACCTGGTGGCC	180
Qy	181	CTGACACGCGCGACCGCGCGGCGATCGAGGATCCGGGAGCGGACTTCCAGTCTTC	240
Db	181	CTGACACGCGCGACCGCGCGGCGATCGAGGATCCGGGAGCGGACTTCCAGTCTTC	240

241 CAGCAGGCGCGCGCGCGCGCTGGCGGACCTTCCGGGCTTCTCTGTCGCGGCTG 300
 Db
 2614 CAGCAGGCGCGCGCGCGCTGGCGGACCTTCCGGGCTTCTCTGTCGCGGCTG 2673
 QY
 301 CAGACCTTACAGACATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
 Db
 2674 CAGACCTTACAGACATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2733
 QY
 361 AGGACAGAGTGTCTTCCCGACGCTGGGAGGCTTATTCTCGGGCTCCGAGGCGAGCTG 420
 Db
 2734 AAGGACAGAGTGTCTTCCCGACGCTGGGAGGCTTATTCTCGGGCTCCGAGGCGAGCTG 2793
 QY
 421 AAGCGCGCGCGCGCGCGCGCTTCTTTCGAGGCGAGAGTCTCTGAGGCTCTGAGGCTCCGCTG 480
 Db
 2794 AAGCGCGCGCGCGCGCGCGCTTCTTTCGAGGCGAGAGTCTCTGAGGCTCTGAGGCTCCGCTG 2853
 QY
 481 CCGCGGAGAGAGTGTGGCAGCGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 540
 Db
 2854 CCGCGGAGAGAGTGTGGCAGCGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 2913
 QY
 541 TGCAGAGCTGGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
 Db
 2914 TGTGAGAGCTGGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2973
 QY
 601 GGCAGGCTGTGGAGCAGGAGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 Db
 2974 GGCAGGCTGTGGAGCAGGAGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 3033
 QY
 561 GAGAACAGCTCATGCTCTTCTCAAGTAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
 Db
 3034 GAGAACAGCTCATGCTCTTCTCAAGTAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3093
 QY
 721 GAGGGGCGCGCGCGGAGGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756
 Db
 3094 ACCGGCGCTGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3153
 QY
 757 -GGGCGCTGGCGGAGGCTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
 Db
 3154 CAGAACCTGCTGCGGAGGCTTCTGCTATAGTTCACGTTTCTGATGATGCGGAGG 3213
 QY
 814 AAGGAGGCGCGGAGGAG 829
 Db
 3214 AAGGAGGCGCGGAGGAG 3229

RESULT 6

US-09-880-107-2178
 ; Sequence 2178, Application US/09880107
 ; Patent No. US20020142981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horne, Darci T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scherf, Owe
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-WO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2178
 ; LENGTH: 3394
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L29548
 ; US-09-880-107-2178

Query Match

Best Local Similarity 59.9%; Score 496.8; DB 9; Length 3394;
 Matches 560; Conservative 77.1%; Pred No. 4.2e-108; Mismatches 157; Indels 39; Gaps 3;
 QY 1 CCCTGGGGGCGAGATGATCATCTTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
 Db 1378 CCCTGGGGGCGAGATGATCATCTTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437
 QY 61 CCGGGGCG 120
 Db 1438 CCGGGGCG 1493
 QY 121 CCGGTCCACACCG 180
 Db 1494 -----CCACCG 1545
 QY 181 CTGAACAGCG 240
 Db 1546 CTGAACAGCG 1605
 QY 241 CAGCAGGCG 300
 Db 1606 CAGCAGGCG 1665
 QY 301 CAGGACCTTACAGCAGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
 Db 1666 CAGGACCTTACAGCAGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1725
 QY 361 AGGACAGAGTGTCTTCCCGACGCTGGGAGGCTTATTCTCGGGCTCCGAGGCGAGCTG 420
 Db 1726 AAGGACAGAGTGTCTTCCCGACGCTGGGAGGCTTATTCTCGGGCTCCGAGGCGAGCTG 1785
 QY 421 AAGCGCGCGCGCGCGCGCGCTTCTTTCGAGGCGAGAGTGTCTCTGAGCAGCGCGCGCG 480
 Db 1786 AAGCGCGCGCGCGCGCGCTTCTTTCGAGGCGAGAGTGTCTCTGAGGCGAGCGCGCGCG 1845
 QY 481 CCGCGGAGAGGCTGTGGCAGCGCTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
 Db 1846 CCGCGGAGAGGCTGTGGCAGCGCTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1905
 QY 541 TSCGAGACGCTGGCGGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
 Db 1906 TGTGAGAGCTGGCGGAGCGAGGCTTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1965
 QY 601 GCGAGGCTGTGGAGCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 Db 1966 GCGAGGCTGTGGAGCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2025
 QY 661 GAGAACAGCGCTCATGACCTTCTTCCAAAGTAGGCGCGCGCGCGCGCGCGCGCGCGCG 720
 Db 2026 GAGAACAGCGCTCATGACCTTCTTCCAAAGTAGGCGCGCGCGCGCGCGCGCGCGCGCG 2085
 QY 721 GAGGGGCGCGCGCGGAGGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756
 Db 2086 ACCGGCGCTCGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2145
 QY 757 -GGGCGCTGGCGGAGGCTTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
 Db 2146 CAGGACCTGGCTGCCATCTTCTCTGTATAGTTCACGTTTCTGATGATGATGATGAT 2205
 QY 814 AAGGAGGCGCGGAGGAG 829
 Db 2206 AAGGAGGCGCGGAGGAG 2221

RESULT 7

US-10-060-036-144
 ; Sequence 144, Application US/10060036
 ; Publication No. US20030073144A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Benson, Darin R.
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Lodes, Michael J.

; GENERAL INFORMATION:
 ; APPLICANT: Campochiaro, Peter A.
 ; APPLICANT: Dixon, Katharine H.
 ; APPLICANT: Brazzell, Romulus K.
 ; TITLE OF INVENTION: METHOD FOR TREATING OCULAR
 ; TITLE OF INVENTION: NEOVASCULARIZATION
 ; FILE REFERENCE: 4-31881A
 ; CURRENT APPLICATION NUMBER: US/10/080,797
 ; CURRENT FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 551
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-080-797-2

Query March 50.3%; Score 417; DB 13; Length 551;
 Best Local Similarity 85.3%; Pred. No. 3.1e-89;
 Matches 465; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY	140	ACACCCACGAGACTTCAGCTGGTGTGCACTGGTGGCCCTGAAACGCGCAGCGG	199
Db	1	ACAGCCACCGGAGTCTCCAGCCGGTGTCTCACCTGGTTCGCTCAACAGCCCCCTGTG	60
QY	200	CGGGATGCGAGGATCCGGGAGGAGACTCCAGTGCCTCCAGCAGCGCGCGCGCGG	259
Db	61	CGGGATGCGGGGATCCGGGGGCGGACTTCCAGTGGTTCAGCAGCGCGCGCGCGG	120
QY	260	GGCTGGCGCGCACCTTCGGGACCTTCTGTGCTGGCGCTGCAGGACCTCTACAGCATG	319
Db	121	GGCTGGCGGGACCTTCGGCGCTTCTGTCTCTCGCGCTGCAGGACCTGTACAGCATG	180
QY	320	TGCGCGCGCGCGACCGCAGCGGGGTGCCGTGCTCACTCAGGACGAGGTGCTCTTCC	379
Db	181	TGCGCGGTGCGGACCGCGCAGCGGTGCCATCGTCAACCTCAGGACGAGTGTCTTTC	240
QY	380	CAGTGGGAGGCGCTTATCTCGGCTCCGAGGCGCAGTGAAGCCCGGGCCCGCATCT	439
Db	241	CAAGTGGGAGGCTCTGTCTCAGGCTCTGAGGTCGCGTGAAGCCCGGGCACGCACT	300
QY	440	TCTCTTTTCGAGCGCAGAGATGCTCTGACGACCCCGCTGCGCCCGGAGAGCGTGTGC	499
Db	301	TCTCTTTTCGAGCAAGGAGCTCTGAGGACACCCACCTGGCCCCCAGAGAGCGTGTGC	360
QY	500	ACGGCTTCGACCCCGAGCGGGCGCGCTGACCGACAGCTACTGCGAGACGTGCGCGG	559
Db	361	ATGGCTTCGACCCCGAACCGGGCGGAGGCTGACCGAGAGCTACTGTGAGACGTGCGG	420
QY	560	AGGCCCCCGCGCCACCGGGCAGGCGTCTGCTGCTCGCGGCGAGGCTGTGAGAGCG	619
Db	421	AGGCTCCCTCGGCGCACCGGCGCAGGCTCTCTGCTGCTGGGGGCGAGGCTCTCTGGG	480
QY	620	AGGCGCGAGCTCGCCGACCGCTCTGTTGGTCTCTGATCGAGAACAGCGTCATGACCT	679
Db	481	GTGCGCGAGTGGCCATCAGCGCTACATCGTGTCTGTGATTGAGAACAGCTTCATGATG	540
QY	680	CCTTC 684	
Db	541	CCTCC 545	

RESULT 13
 US-10-131-241-51
 ; Sequence 51, Application US/10131241
 ; Publication No. US20030012792A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Holaday, John W.
 ; APPLICANT: Fortier, Anne H.
 ; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
 ; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
 ; FILE REFERENCE: 05213-0344 43170-271565
 ; CURRENT APPLICATION NUMBER: US/10/131,241

RESULT 14
US-10-210-172-161
; Sequence 161, Application US/10210172
; Publication No. US20040094328A1
; GENERAL INFORMATION:
; APPLICANT: Kektud, Rameeh
; APPLICANT: Maller, Charles
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol

Db	1	GACTTCAGCGGTGCTCTCACTGGTGGCTCAAAGCCCCCTGTCAAGCGGCATCGG	60
Qy	211	GGCATCCGGGAGCGGACATTCACGTGCTTCACAGCGCGCGCGGGGTGGCCGCG	270
Db	61	GGCATCCCGGGGCGGACATTCACGTGCTTCACAGCGCGCGCGGGGTGGCGGC	120
Qy	271	AGCTTCCGGGCTTCTCTGTGCTGCGGCTCAGGACCTCTACAGCATCGTGGCGCCGCGC	330
Db	121	ACCTTCCGGCTTCTCTGTGCTGCGGCTCAGGACCTGTACAGCATCTGCGCGCGTGC	180
Qy	331	GACCCACCGGGGTCCGTCGTCACCTCAGGAGGAGGTGCTCTTCCCACGCTGGGAG	390
Db	181	GACCGCGCAGCGGTGCCCATCGTCAACCTCAAGGACAGCTGCTGTTTCCCAGCTGGAG	240
Qy	391	GCCTTATTCTCGGCTCCGAGGCGCAGCTGAAGCCCGGGCCCGCANCTTCTTTTCAC	450
Db	241	GCTCTGTTCTCAGGCTCTGAGGGTCGCGCTGAAGCCCGGGCGCAGCATCTTCTCTTTCAC	300
Qy	451	GGCAGAGATGTCCTGACGACCCCGCTGCGCCCGGAGAGGCTGTGGCAGCGCTCCGAC	510
Db	301	GGCAAGGACGCTCTGAGGACCCCACTGCGCCCGCAGAGAGCTGTGGCATGCGCTCGAC	360
Qy	511	CCACAGCGGCGCGCTGAACGACAGTACTGCGAGACGTGGCGGACGAGGCCCGCGG	570
Db	361	CCCAACGGGCGCAGCTGACCGAGAGCTACTGTGAGACGCTGGCGGACGAGGCTCCCTCG	420
Qy	571	GCACCGGGCAGGGGTGCTGCTGCTCGCGGCGAGGCTGTGGAGCAGGAGCCCGCAGC	630
Db	421	GCACGGGCGAGGCGCTCTCTGCTGCTGGGGGACAGGCTCTTGGGCGAGTGCCTGGGAGC	480
Qy	631	TGCGCGCCACGCTCTGTTGGTCTCTGTCATCGAGAACAGCGGTGATGACCTTCCTTC	684
Db	481	TGCCATACGCGCTACATCGTCTCTGATTGTAGAACAGCTTCACTAGTACGCTCC	534

```

RESULT 15
US 10-042-347-6
; Sequence 6, Application US/10042347
; Reference 6, N US20030114370A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fragments

```

Search completed: March 29, 2004, 16:40:25
Job time : 340.85 secs

Query Match 49.6%; Score 410.8; DB 14; Length 534;
Best Local Similarity 85.6%; Pred. No. 9e-88;
Matches 457; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 151 GACTTCAGCTGTTGTGCACTGTGTGCGCCCTGAACAGCCCGAGCGGGCGGCATGCCGA 210

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 11:07:01 ; Search time 2546.3 Seconds

Sequence: 829
(without alignments)
9722.244 Million cell updates/sec

Title: US-09-938-391-1

Perfect score: 1 cctggcgccgagatgacat.....ataaaaggaagccaagag 829

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pig:*

27: em_gss_vri:*

28: gb_gse1:*

29: gb_gse2:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	503.2	60.7	4230	11	BC063833 Homo sapi
2	493.2	59.5	881	14	CD105862 AGENCOURT
3	459.6	55.4	929	13	BQ672290 AGENCOURT
4	451	54.4	979	13	BQ673186 AGENCOURT

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	445.2	53.7	551	10	BF074459
6	441.6	53.3	832	12	BG387051
7	436.8	52.7	4192	11	BC062931 Mus muscu
8	428	51.6	843	10	BF385854
9	425.8	51.4	715	9	AU125614
10	425.8	51.4	947	13	BUS56872 AGENCOURT
11	422.8	51.0	707	10	BE908201
12	422.4	51.0	757	10	BE906253 601500458
13	417	50.3	874	12	BI412588
14	413	49.8	504	14	CF789984 857322 MA
15	401.6	48.4	657	14	CB444165 695295 MA
16	398.2	48.0	944	13	BUS59398 AGENCOURT
17	394.4	47.6	703	13	BUE15520 UI-H-PGO-
18	387	46.7	683	12	BM683067 UI-E-EO1-
19	386.6	46.6	682	9	AM085983 x20f03 X
20	385.6	46.5	735	14	CF728236 UI-M-HB0-
21	379.6	45.8	618	9	AV696242 AV696242
22	379.2	45.7	715	14	CB596713 AGENCOURT
23	378.2	45.6	897	12	BI080524 602877005
24	376.8	45.5	611	10	AW911243 ur-83h10.Y
25	376.6	45.4	782	9	AI326391 mm18f09.X
26	374.4	45.2	884	12	BI161007
27	373.8	45.1	753	12	BI904605 603168411
28	373.4	45.0	835	12	BI526580 602925454
29	368.6	44.5	723	12	BI247582 602960041
30	366.6	44.2	1093	13	BQ723254 AGENCOURT
31	364.6	44.0	668	14	CF724654 UI-M-GZ0-
32	356.8	43.0	915	10	BF166139 601776586
33	356.6	43.0	664	13	BUE32049 UI-H-FE1-
34	354.2	42.7	650	9	AI858615 w140f01.X
35	353.8	42.7	846	13	BUS40812 AGENCOURT
36	353	42.6	720	12	BI147444 602914008
37	352.6	42.5	639	12	BM998137 UI-H-DT1-
38	350	42.2	634	13	BQ772348 UI-H-EZ1-
39	350	42.2	660	10	AW192502
40	349.4	42.1	652	13	BUS32506
41	348.8	42.1	747	12	BG967333 602833649
42	348.6	42.1	587	9	AA288198 vb15b10.X
43	348.2	42.0	634	9	AI970297 wr09c02.X
44	343.8	41.5	745	10	AW243446 xm95d11.X
45	343.6	41.4	690	12	BI213999 602936756

ALIGNMENTS

RESULT 1	BC063833	4230 bp	mRNA	linear	HTC 09-DEC-2003
BC063833	Homo sapiens cDNA clone IMAGE:6181818, containing frame-shift errors.				
LOCUS	BC063833				
DEFINITION	BC063833.1	GI:39645297			
ACCESSION	BC063833				
VERSION	BC063833.1				
KEYWORDS	HTC				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 4230)				
AUTHORS	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Hopsch S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh P., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Wokley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,				

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywicki, M.I., Skalek, U., Smalhus, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 1247932
 2 (bases 1 to 4230)
 Direct Submission
 Submitted (08-DEC-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywicki, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Nees, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Scott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 141 Row: c Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 18765747
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1. 4230
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6181818"
 /tissue_type="peripheral Nervous System, dorsal root
 ganglion"
 /clone_lib="Lupski_dorsal_root_ganglion"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 60.7%; Score 503.2; DB 11; Length 4230;
 Best Local Similarity 77.6%; Pred. No. 3.8e-63;
 Matches 664; Conservative 0; Mismatches 153; Indels 39; Gaps 3;
 QY 1 CCTGGGGGCGAGATGACATCTCTGGCGGGCCCGCGCGCTGTGGAGCCCGACCCCTAC 60
 Db 3369 CCTGGGGGCGAGATGACATCTCTGGCGGGCCCGCGCGCTGTGGAGCCCGACCCCTAC 3428
 QY 61 CCGGGGCGCCCGACACCGCTCTACGTGCACTTCCAGCGCGGTCTGCGCCCACTGTGGG 120
 Db 3429 CCGGGGCGCCCGACACCGCTCTACGTGCACTTCCAGCGCGGTCTGCGCCCACTGTGGG 3484
 QY 121 CCGGTCCACACCCACACCCACACCCACAGACTTCCAGCTGTGTGTGTGTGTGTGTGTGT 180
 Db 3485 -----CCACCCCGCCCGACAGCCAGCCAGCTTCCAGCGCGGTCTGCTCCACTGTGTGG 3536
 QY 181 CTGAACAGCCCGGAGCGGGCGCATGCCAGGACATCCCGGGAGCGGACTTCCAGTGTTC 240
 Db 3537 CTGAACAGCCCGGAGCGGGCGCATGCCAGGACATCCCGGGAGCGGACTTCCAGTGTTC 3596
 QY 241 CAGCAGCGCGCGCGCGGGGTGGCGCGGACACTTTCGCGGCGCTTCTGTGTGTGTGTGTGT 300

Db 3597 CAGCAGCGCGCGCGCGGGCTGTGGCGGCGACCTTCCGCGGCTTCTGTGTGTGTGTGTGTGT 3656
 QY 301 CAGGACCTCTACAGCATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
 Db 3657 CAGGACCTCTACAGCATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3716
 QY 361 AGGACCAAGGCTCTTTCCTCCAGCTGGAGGCTTATTTCTCGGGTCTCGAGGCGCAGCTG 420
 Db 3717 AAGGACGAGCTGTGTTTCCAGCTGGAGGCTCTGTTCTCAGGCTCTCAGGCTCTGAGG 3776
 QY 421 AAGCCCGGGCGCGCATCTTCTTTCGACGCGCAGAGATGCTCTGCAGCACCCCGCGCTGG 480
 Db 3777 AAGCCCGGGCGCGCATCTTCTTTCGACGCGCAGAGAGCTCTGAGGACCCCGACCTGG 3836
 QY 481 CCCCGGAAGAGCTGTGGCAGCGCTCCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 540
 Db 3837 CCCCGGAAGAGCTGTGGCATGTGCTCGGACCCCGCGCGCGCGCGCGCGCGCGCGCGCTAC 3896
 QY 541 TGGCAGACGCTGGCGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
 Db 3897 TGTGAGACGCTGGCGGCGAGGCTCTCTCGGACCGGCGCGCGCGCGCGCGCGCGCGCG 3956
 QY 601 GCGAGGCTGTGGAGCAGGAGCGCGCGAGCTCCGCGCCACGCGCTTCTGTGTGTGTGTGTGTGT 660
 Db 3957 GCGAGGCTCTGGCGGCGAGAGTGGCGCGAGCTGCGCATCAGCGCTACATCTGTCTGCAAT 4016
 QY 661 GAGAACAGCGCTGATGCTCTTCTCCAAAGTAGGGCGCGCGCGCGCGCGCGCGCGCGCG 720
 Db 4017 GAGAACAGCTTCATGCTGCTCCAAAGTAGGACCGCGCGCTGATGAGATGCGCGGAGG 4076
 QY 721 GAGGGGCGCGCGCGAGAGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756
 Db 4077 ACCGGCGCTCGGAGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4136
 QY 757 --GGGCGTGGCGGCGAGCTTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
 Db 4137 CAGGACCTGTGCTGCCATCTTCTCTGTATGATTCAGTTTCATGTAATCTCAAGAAATA 4196
 QY 814 AAGGAAGCGCAAGAG 829
 Db 4197 AAGGAAGCGCAAGAG 4212
 RESULT 2
 LOCUS CD105862 881 bp mRNA linear EST 15-MAY-2003
 DEFINITION AGENCOURT 14021798 NIH_MGC_179 Homo sapiens cDNA clone
 IMAGE:130365831 5', mRNA sequence.
 ACCESSION CD105862
 VERSION CD105862.1 GI:30759036
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 881)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Preparation: Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM424 row: g column: 24
 High quality sequence stop: 689.
 Location/Qualifiers
 1. 881

FEATURES

source

ORIGIN

Query Match 59.5%; Score 493.2; DB 14; Length 881;
Best Local Similarity 83.9%; Pred. No. 8.9e-62;
Matches 574; Conservative 0; Mismatches 98; Indels 12; Gaps 1;

1 CCTTGGGGGACATGATGATCTCTGGGGGCCCCCGGCTGCTGACCCCGAGCCCTAC 60
166 CCTTGGGGGACATGATGATCTCTGGGGGCCCCCGGCTGCTGACCCCGAGCCCTAC 225

61 CCGGGGCCCCGACACACACCGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 120
226 CCGGAGCCCCGACACACACCGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 281

121 CCGGTCCACACACACACACCGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 180
282 -----CCACCCGCGCACACCGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 333

181 CTGAACCGCGGACGCGGCGGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 240
334 CTGAACCGCGGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 393

241 CAGCAGCGGCGGCGGCGGCTGCGCCACTGCTGGCGGCTTCTGCTGCTGCGGCTG 300
394 CAGCAGCGGCGGCGGCGGCTGCGCCACTGCTGGCGGCTTCTGCTGCTGCGGCTG 453

301 CAGGACCTCTAGCAGCTGCTGCGGCGGCTGCGCCACTGCTGGCGGCTTCTGCTGCTGCGGCTG 360
454 CAGGACCTCTAGCAGCTGCTGCGGCGGCTGCGCCACTGCTGGCGGCTTCTGCTGCTGCGGCTG 513

361 AGGACGAGGTGCTTCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 420
514 AAGGACGAGGTGCTTCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 573

421 AAGCGGCGGCGGCGGCTTCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 480
574 AAGCGGCGGCGGCGGCTTCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 633

481 CCGCGGAGGAGCGGTGCGGCGGCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 540
634 CCGCGGAGGAGCGGTGCGGCGGCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 693

541 TCGGAGAGGTGCGGCGGAGCGGCGGCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 600
694 TGTGAGAGGTGCGGCGGAGCGGCGGCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 753

601 GCGAGGTGCTGAGGAGGCGGCGGCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 660
754 GCGAGGTGCTGAGGAGGCGGCGGCTTCCCGAGCGGCTTCTGCTGCTGCGGCTTCTGCTGCTGCGGCTG 813

661 GAGAACGCGGTGATGACCTCTCTTC 684
814 GAGAACGCGGTGATGACCTCTCTTC 837

RESULT 3
BQ672290 929 bp mRNA linear EST 15-JUL-2002
LOCUS AGNCOURT 8354946 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275261
DEFINITION 5', mRNA sequence.

ACCESSION BQ672290
VERSION BQ672290.1 GI:21783124
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 929)
NIH-MGC <http://mgi.nci.nih.gov/>
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM2458 row: C column: 06
High quality sequence stop: 588.
Location/Qualifiers
1..929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6275261"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_102"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 55.4%; Score 459.6; DB 13; Length 929;
Best Local Similarity 77.4%; Pred. No. 6.4e-57;
Matches 602; Conservative 0; Mismatches 159; Indels 17; Gaps 3;

Qy 1 CCTTGGGGGACATGATGATCTCTGGGGGCCCCCGGCTGCTGACCCCGAGCCCTAC 60
Db 101 CCTTGGGGGACATGATGATCTCTGGGGGCCCCCGGCTGCTGACCCCGAGCCCTAC 160

Qy 61 CCGGGGCCCCGACACACCGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 120
Db 161 CCGGAGGCCCCGACACACCGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 216

Qy 121 CCGGTCCACACACACACCGCTCTAGTGTCTTCCAGCGGCTGCGCCACTGCTGG 180
Db 217 -----CCACCCGCGCACACCGCTCTAGTGTCTTCCAGCGGCTGCTGCGGCTG 268

Qy 181 CTGAACCGCGGACGCGGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCAGTGTTC 240
Db 269 CTGAACCGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCAGTGTTC 328

Qy 241 CAGCAGCGGCGGCGGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTG 300
Db 329 CAGCAGCGGCGGCGGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTG 388

Qy 301 CAGGACCTCTAGCAGCTGCTGCGGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCAGCTTC 360
Db 389 CAGGACCTCTAGCAGCTGCTGCGGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCAGCTTC 448

Qy 361 AGGACGAGGTGCTTCTTCCCGAGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCAGCTTC 420
Db 449 AAGGACGAGGTGCTTCTTCCCGAGCGGCTTCCCGAGCGGCTTCCCGAGCGGCTTCCAGCTTC 508

QY 421 AAGCCCGGGGCGGCATCTTCTTTTCGAGCGCAGAGATGTCTCTGACACCCCGCCTGG 480
 Db 509 AAGCCCGGGGCGGCATCTTCTTTTCGAGCGCAGAGATGTCTCTGAGCAGCAGCCTGG 568
 QY 481 CCCCGAGAGAGCGTGTGGCAGCGCTCGACCCAGCGGCGCGCTCTACCGACAGCTAC 540
 Db 569 CCCCGAGAGAGCGTGTGGCAGCGCTCGACCCAGCGGCGCGCTCTACCGAGAGTAC 628
 QY 541 TGCAGAGCGTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
 Db 629 TGTGAGAGCGTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688
 QY 601 GGCAGG---CTGCTGAGCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
 Db 689 GGGCAGGCTCTTGGGCGCAGAGTGTCCCGGAGAGTGTCCCATTCAGCGCTCATCTG 748
 QY 658 ATCGAGAACAGCGTGTGACCTCTTCTTCTCAAG--TAGGGCGCGCGCGCGCGCGCG 715
 Db 749 ATTGAGAACCGCTTCTGATGCTGCTCAAGTAGTACCGCGCTGTGGCAATGCGAGGG 808
 QY 716 CGGGGAGGGGGCG 773
 Db 809 AAAAGGCGCGCGCACTCTCGAGAAACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 866

RESULT 4
 BQ673186
 LOCUS
 DEFINITION
 AGENCOURT_8414390 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6272287
 5', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 NIH-MGC <http://imgc.nhl.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM2450 row: 9 column: 08
 High quality sequence stop: 638.

FEATURES
 source
 Location/Qualifiers
 1..979
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6272287"
 /issue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_102"
 /note="Organ: salivary gland; vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dt priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-cDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 54.4%; Score 451; DB 13; Length 979;
 Best Local Similarity 78.8%; Pred. No. 1.1e-55;

Matches 606; Conservative 0; Mismatches 145; Indels 18; Gaps 5;
 QY 1 CCCTGGGGGGGAGATGATCATCTCTGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 60
 Db 101 CCCTGGGGGGGAGATGATCATCTCTGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 160
 QY 61 CCCGGGGCGCGGACCAACGCGCTCTACGTGTCACTTCCAGCCCGGCTCGGCCCACTGTGG 120
 Db 161 CCCGGAGCGCGCGACCAACGAGCTCTAGTGTCACTTCCAGCCCGGCTCGGCCCACTGTGG 216
 QY 121 CCCGTCACACACCCACACCCACACACCGAGGACTTCCAGCTGTGTCTGACCTGTGGCC 180
 Db 217 -----CCACCCCGCCACAGCCACCGGACTTCCAGCCCGGCTCTCCACTGTGTGG 268
 QY 181 CTGAACAGCCCGGAGCGCGCGCATGCGAGGCAATCCGGGGAGCGGACTTCCAGTGTTC 240
 Db 269 CTCAACAGCCCGCTGTGAGCGGCAATGCGGGGCAATCCGGGGAGCGGACTTCCAGTGTTC 328
 QY 241 CAGCAGCG 300
 Db 329 CAGCAGCG 388
 QY 301 CAGGACCTTACAGCATCTCTCTTTTCGAGCGCAGAGATGTCTCTGACACCCCGCCTGG 360
 Db 389 CAGGACCTGTACAGCATCTCTCTTTTAAACGGCAAGGACGCTCTTGAGGACCCCGCCTGG 448
 QY 361 AGGAGCAGGTGCTCTCTCCAGCTGGGAGGCTTATTCTGGGGCTCGAGGGCGAGCTG 420
 Db 449 AAGACAGCTGTGTCTTCCAGCTGGGAGGCTCTGTCTCAGGCTCTGAGGGTCCGCTG 508
 QY 421 AAGCCCGGGGCGCGCATCTCTCTTTTCGAGCGCAGAGATGTCTCTGACACCCCGCCTGG 480
 Db 509 AAGCCCGGGGCGCGCATCTCTCTTTTAAACGGCAAGGACGCTCTTGAGGACCCCGCCTGG 568
 QY 481 CCCCGGAGAGCGTGTGGCAGCGCTCCGACCCCGCGCGCGCGCGCGCGCGCGCGCG 540
 Db 569 CCCCGAGAGAGCGTGTGGCAGCGCTCCGACCCCGCGCGCGCGCGCGCGCGCGCGCG 628
 QY 541 TGGCAGAGCTGTGGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 599
 Db 629 TGTGAGAGCTGTGGCGAGGAGCTCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 688
 QY 600 GGCAGGCTCTGAGCAGGA--GGCGCGAGCTGCGCGCGCGCTT--CGTGGTGTCT 655
 Db 689 GGGCAGGCTCTTGGGGCAGAGTGGCGCGAGTGGCGAGTGGCATCGCGCTCTACATCGTGGCTCT 748
 QY 656 GCATCGAGAACGCGTCTATGACCTCTCTTCTCAAGTAGGGCGCGCGCGCGCGCGCGCAGG 715
 Db 749 GCATTGAGAACAGCTCTATGACTGCTCTC--CAAGAAACACCGCGCTGGAAATCGGAATGG 807
 QY 716 CGGGGAGGGGGCTGG 764
 Db 808 CCCGAAAGAGACCGGGCGGCTCGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 856

RESULT 5
 BQ704459
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Bos taurus (cow)
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 551)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.B., White,J., Cho,J., Fahrenkrug,S.C.,
 Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
 Chitko-McKown,C.G., Pettea,G., Holt,I., Karamycheva,S., Liang,F.,
 Quackenbush,J. and Keele,J.W.

BF074459 551 bp mRNA linear EST 25-APR-2001
 221883 MARC 280V Bos taurus cDNA 5', mRNA sequence.
 BQ704459
 BF074459.1 GI:10867970
 EST.
 Bos taurus (cow)
 Bos taurus

REFERENCE
 AUTHORS
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.B., White,J., Cho,J., Fahrenkrug,S.C.,
 Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
 Chitko-McKown,C.G., Pettea,G., Holt,I., Karamycheva,S., Liang,F.,
 Quackenbush,J. and Keele,J.W.

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Sequence evaluation of four pooled-tissue normalized bovine cdna
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
11282978

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTACACGACG
Plate: 81 row: F column: 9
Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers
1. .551
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="MARC 2BOV"
/notes="Vector: pCMV SPOT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semiteadonous muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

FEATURES
source

1. .551
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="MARC 2BOV"
/notes="Vector: pCMV SPOT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semiteadonous muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN

Query Match 53.7%; Score 445.2; DB 10; Length 551;
Best Local Similarity 88.5%; Pred. No. 7.2e-55;
Matches 483; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 163 GTGCTGACCTGTGGCCCTGAAACAGCCGCGAGCGGGGCGGATGCGAGGATCCCGGGA 222
Db 1 GTGCTGACCTGTGGCGCTTCAACAGCCGCGAGTGGCGGGCGCTGGCGGATCCCGGCG 60

Qy 223 GCGGACTTCCAGTGTCTCCAGCAGGCGCGCGCGGGGCTGGCGGACCTTCGCGGCC 282
Db 61 GCGGACTTTCAGTGTCTCCAGCAGGCGCGCGCGGGCTGGCGGACCTTCGCGCG 120

Qy 283 TTCTGTGCTGCGGCTGACGACCTCTACAGATCGTGGCGCGCGCGAGCGGACCGG 342
Db 121 TTCTGTGCTTCCGCGGTTCAGGACCTGTACAGATCGTGGCGCGCGCGAGCGTCCACC 180

Qy 343 GTGCGGTGCTCACTCAGGACGAGGTGCTCTCCAGCTGGAGGCTTATTCG 402
Db 181 CTGCGGTGCTCACTCAGGACGAGGTGCTCTCTAGCTGGAGGCTTATTCG 240

Qy 403 GGCTCCAGGCGCAGCTGAAGCCGCGGCGCGCGATCTTCTTTCCAGGCGAGAGATGC 462
Db 241 GGCTCCAGGCGCAGCTGAAGCCGCGGCGCGCGATCTTCTCTTCCAGGCGAGAGATGC 300

Qy 463 CTGAGACACCCGCTGGCGCGGAGAGGCTGTGGCAGGCTCCGACCCAGCGGCGG 522
Db 301 CTTTACGATCCCACTGGCGCGGAGAGAGCGTGTGCGACGGCTCTATCCCAAGCGGCGG 360

Qy 523 CGCTGACCCACAGCTACTCCGAGAGCTGGCGGAGAGGCGCGCGGCGGCGGCGGAG 582
Db 361 CGCTGACCCAGAGCTACTCCGAGAGCTGGCGGAGAGGCGGCGGCGGCGGCGGAG 420

Qy 583 GCGTGTGCTGTGGCGGCGGAGCTGTGGAGCAGGAGCGCGAGCTGCGCGGCGGCGG 642
Db 421 GCGTGTGCTGTGGCGGCGGCGGCTGTGGAGCAGGAGCGGCGGCGGCGGCGGCGG 480

Qy 643 TTGCTGTGCTGTGCTGACGAGACAGCTCATGACCTCTCTTCCAAAGTAGGGCGCGG 702
Db 481 TTCAATTGCTCTGTGATCGAGAAACAGCTCATGACCTCTCTCTCCAAAGTAGGGCGGCTGCG 540

Qy 703 GCCCAC 708
Db 541 GGGCAC 546

RESULT 6
BG387051
LOCUS
DEFINITION
602454749F1 NIH_MGC_15 Homo sapiens cdna clone IMAGE:4582933 5',
mRNA sequence.
BG387051
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 832)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cdna Library Preparation: Ling Hong/Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1306 row: a column: 14
High quality sequence stop: 679.
Location/Qualifiers
1. .832
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4582933"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH108 (phage-resistant)"
/clone_lib="NIH MGC_15"
/note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cdna made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)";

ORIGIN

Query Match 53.3%; Score 441.6; DB 12; Length 832;
Best Local Similarity 82.1%; Pred. No. 2.5e-54;
Matches 550; Conservative 0; Mismatches 104; Indels 16; Gaps 3;

Qy 1 CCCTGCGGCGGAGATGACATCTCTGGCGCGGCGCGCGCGCTGCTGGAGCCCGAGCCCTAC 60
Db 89 CCCTGCGGCGGAGATGACATCTCTGGCGCGGCGCGCGCTGCTGGAGCCCGAGCCCTAC 148

Qy 61 CCGCGGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 149 CCGCGGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 203

Qy 121 CCGGTCACACCCACACCCACACCCACACCCAGGAGCTTCCAGCTGGTGTCTGACCTGGTGCC 180
Db 204 -----CCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 255

Qy 181 CTGAACAGCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 256 CTCAACAGCCCGCTGTTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 315

Qy 241 CAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300


```
QY 538 TACTGCGAGACGTGGGGGACGAGGCGCCCGGGGCGCCACCGGGCAGCGCTGCTGCTGCTG 597
Db 3841 TACTGTGAGACATGGCGAATCTACTTGGGCTACAGGTTCAGGCTTCTCCCTGCTG 3900
QY 598 GCGGGCGAGTGTGAGACGAGGAGCGCGAGCTGCGCGCAGCCCTTGGTGTGCTTGC 657
Db 3901 TCAGGCGAGGCTCTGGAACAGAAAGTGGAGCTGCCAACAGCTACATGCTGCTGTC 3960
QY 658 ATCGAGAACAGCTCATGACCTCTTCTCCAAAGTAGGGCGCGCGCCCGCAGCGAGCGG 717
Db 3961 ATTGAGAAATAGCTTCATGACCTCTTCTCCAAATAGGGCTCTGCCAGCTAGGTGGCGAG 4020
QY 718 GGGGAGS-----GGCGCGCGCAGGAGGATCGCGCGCGCGCGCGCGCGCGCGCG 761
Db 4021 ACAGAGCCCATGAGAACTTTGACACAGCGCAGGAGCAATCAGTCAGCAGCCAGCGGCTC 4080
QY 762 TGGCGGCGAGCGCTGCTGCTGCGACCGT--CACGTTTAAATGTAATCTCAAGAAATAAAGGA 819
Db 4081 TGGCTGGGATACAACTCTGTATAGTTCCCATTTTATGTTAACTTCAAGAAATAAAGGA 4140
QY 820 AGCCAAAGAG 829
Db 4141 AGCCAAAGAG 4150

RESULT 8
BF385854
LOCUS
DEFINITION 602046021F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4195660 5',
mRNA sequence.
ACCESSION BF385854
VERSION BF385854.1 GI:11367159
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-1@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9531 row: i column: 05
High quality sequence stop: 761.
FEATURES
Location/Qualifiers
1..843
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/cdones="IMAGE:4195660"
/lib_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_L19"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Query Match 51.6%; Score 428; DB 10; Length 843;
Best Local Similarity 77.9%; Pred.No. 2.3e-52;
Matches 547; Conservative 0; Mismatches 140; Indels 15; Gaps 2;
QY 1 CCCTGGCGGGCAGATGACATCTCTTGGCGCGCGCGCGCGCTGCTGAGCCCGCCAGCCCTAC 60
```

RESULT 9

AUI25614 715 bp mRNA linear EST 01-AUG-2002
AUI25614 NT2RM4 Homo sapiens cDNA clone NT2RM4001897 5', mRNA
sequence.

ACCESSION AUI25614

VERSION AUI25614.1 GI:10950330

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 715)

Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and
Isogai, T.TITLE HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,
Suzuki, Y., Sugano, S., Isogai, T.)

JOURNAL Unpublished (2000)

CONTACT Takao Isogai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3975

Db 551 ACATGGGAACTGAACTACTGGGGCTACAGTTCAGGCTCTCTCCCTGTGTGTCAGGCGG 610
Qy 607 CTGCTGGAGCAGGAGGCGCGAGCTGCGGACGCTTGTGTGTCTGTGCTCTGACGAC 666
Db 611 CTCTTGAACAGAAAGTGGAGCTGCGACACAGCTACATCTCTGTGCTATGAGAT 670
Qy 667 AGCTCATGACCTCTCTCTTCCAAAGTAGGCGCG 699
Db 671 AGCTTCATGACCTCTTCTCTCAATAGGCGCTC 703
RESULT 14
LOCUS CF789984 504 bp mRNA linear EST 21-OCT-2003
DEFINITION 867322 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION CF789984
VERSION CF789984.1 GI:37794545
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 504)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
Nonnenan,D.J., Wray,J.E. and Keele,J.W.
TITLE Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TWK8006 row: B column: 3
Seq primer: GTAATACGACTCCTATAGG.
Location/Qualifiers
FEATURES
source
1..504
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4PIG"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."

ORIGIN
Query Match 49.8%; Score 413; DB 14; Length 504;
Best Local Similarity 89.0%; Pred. No. 3.2e-50;
Matches 446; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
Qy 200 GCGCATGCGGAGGATCCGGGGAGCGGACTTCCAGTGTCTCCAGCAGGCGCGCGCGG 259
Db 1 GCGGCTGCGGGCATCGCGGGCGGACTTCCAGTGTCTCCAGCAGGCGCGCGCGTGG 60
Qy 260 GGTGCGCGGACCTTCGGGGCTTCTGTGTGCGGGCTGCGGACCTCTACAGATCG 319
Db 61 GGTGCGGGGACCTTCGGGGCTTCTGTGTGCGGGCTGCGGACCTCTACAGATCG 120
Qy 320 TGGCGCGCGCGACCGGCGGTGCGGTCTGCTCAACTCTCAGGACGAGGTGCTCTCC 379
Db 121 TGCAGCGCGCGACCGGCGCGGTGCGGTCTGCTCAACTCTCAGGACGAGGTGCTGTTCC 180
Qy 380 CCAGCTGGAGGCTTATTCTCGGGCTTCGAGGGCCAGCTGAAGCCCGGGCGCGCATCT 439
Db 181 GGGCGGGGAGGCTTGTCTTCTCGGGCTCTGAGGCGCCAGCTGAAGCCCGCGCGCATCT 240
Qy 440 TCTCTTCGACGCGAGAGATGCTCTGCGACACCCCGCTGGCCCCCGAAGAGCGTGTGC 499

Db 241 TCTCTTCGACGCGAGAGCTCTTCAGACCCCTGGCCCCCAAGAGCGTGTGCG 300
Qy 500 ACGCTCCGACCCAGCGGGCGCGCTGACGACAGCTACTGCGAGAGTGGCGGACGG 559
Db 301 ACGCTCAGACCCAGCGGGCGCGCTGACGAGAGCTACTGTGAGAGTGGCGACTG 360
Qy 560 AGGCCCCCGCGCGCACCGGCGAGCGCTCTGCTGCTGCGGCGGAGGCTGCTGGAGCAGG 619
Db 361 AGGCCAGGACGCGCACCGGCGAGGCTCTCTGCTGCTGCGGCGGCTGCTGGAGCAA 420
Qy 620 AGGCCGCGAGCTGCGCGACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 679
Db 421 AGGCCGCGGCTGCGCGACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Qy 680 CTTTCTCCAAAGTAGGCGCGG 700
Db 481 CCGCTCCAGTAGGATCTCG 501
CB444165 657 bp mRNA linear EST 25-MAR-2003
DEFINITION 695295 MARC 6BOV Bos taurus cDNA 3', mRNA sequence.
ACCESSION CB444165
VERSION CB444165.1 GI:29233914
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 657)
AUTHORS Smith,T.P.L., Roberts,A.J., Ehternkamp,S.E., Chitko-McKown,C.G.,
Wray,J.E. and Keele,J.W.
TITLE A second set of bovine ESTs from pooled-tissue normalized libraries
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: FQY8051 row: I column: 11
Seq primer: TAGAGGCGACAGTCGAGG.
Location/Qualifiers
FEATURES
source
1..657
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 6BOV"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."

ORIGIN
Query Match 48.4%; Score 401.6; DB 14; Length 657;
Best Local Similarity 81.5%; Pred. No. 1.5e-48;
Matches 464; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
Qy 218 GGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGGGGTGGCGGACCTTCC 277
Db 657 GCGGCGCGGACTTCCAGTGTCTCCAGCAGCGCGCGGGGTGGCGGACATTC 598
Qy 278 GGGCTTCTGTGCTGCGGGCTGCGAGCTCTACAGCATCTGCGCGCGCGCGCGCA 337
Db 597 GCGGCTTCTGTCTGCTGCGGGTTGCGAGCACTGTACAGCATCTGCGCGCGCGCGCG 538

QY	338	CGGGGTGCCGTCGTCAACCTCAGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCCTTAT	397
Db	537	CCACCTTGCCGTCGTCAACCTCAGGACGAGGTGCTCTTCCCTAGCTGGGAGGCCCTTGT	478
QY	398	TCTCGGCTCCAGGGCAGCTGAAGCCCGGGCCCGCATCTTCTCTTTTCGACGGCAGAG	457
Db	477	TCTCAGGCTCCAGGGCCAGCTGAAGCCCGGGCCCGCATCTTCTCTTTTCGACGGCAGAG	418
QY	458	ATGTCTTCAGCACCCCGCTGGCCCGGAAAGAGCGTGTGACGGCTCCGACCCACGCG	517
Db	417	ATGTCTTCAGCATCCACCTGGCCCCAAGAGCGTGTGACGGCTCAGACCCACGCG	358
QY	518	GGCGCCGCTGACCGACAGCTACTCGGAGACGTGGCCGACGGAGGCCCGGGGCCACCG	577
Db	357	GGCGCCGCTGACCGAGAGCTACTCGGAGACGTGGCCGACGGAGGCCCGGGGCCACCG	298
QY	578	GSCAGGCTCGTCTGCTGGGGGACGCTGCTGGAGCAGGAGGCCCGGAGCTGCCGCC	637
Db	297	GCCAGGCTCTCTGCTGCTGGGGGACGCTGCTGGAGCAGGAGGCCCGGAGCTGCCGCC	238
QY	638	AGCCCTTCGTGGTCTCTGCATCGAGAACAGCGTCATGACCTCTTCTTCCAAAGTAGGGCC	697
Db	237	AGCCCTTCATTCTCTNTGATCGAGAACAGCTTATGACCTCTCTCTTCCAAAGTAGGGCC	178
QY	698	GGCGGCCCAACGAGCAGGGGGGAGGGGGCGCCCGCAGGAGCATCCGCCGCCCGGGGG	757
Db	177	TCTGCGGGCACCATGGACACAGGTGGACTGGTGAACAGAGTGGCAGAGAGGAGGCCGCC	118
QY	758	GGCTGCGCGGAGCGCTTGCTTGACCGT	786
Db	117	AGCAGGAGGAGGGAATGCCCCGAGCCT	89

Search completed: March 29, 2004, 14:52:44
 Job time : 2570.3 secs

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:28:41 ; Search time 57.2222 Seconds
(without alignments)
1135.676 Million cell updates/sec

Title: US-09-938-391-2

Perfect score: 1239

Sequence: 1 PWRADDILAGPRLDPPY.....CRHAFVVLCIENSVMTSFK 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	100.0	230	5	AAO17429 Canine pr
2	1019	82.2	275	5	AAU76689 Synthetic
3	1019	82.2	310	5	AAU76688 Human col
4	1019	82.2	682	5	ABP41878 Human ova
5	1019	82.2	1301	2	AAW32296 Human alp
6	1019	82.2	1336	2	AAU08694 Human col
7	1019	82.2	1336	6	ABP96308 Human end
8	1019	82.2	1516	5	ABB83471 Human col
9	1019	82.2	1516	5	ABP68617 Human pan
10	1003	81.0	684	2	AAW26327 Human alp
11	1003	81.0	684	5	AAO17357 Human col
12	982	79.3	684	2	AAU25113 Human alp
13	980.5	79.1	1288	2	AAW26328 Mouse alp
14	976	78.8	1288	2	AAW92297 Mouse alp
15	966	78.0	184	5	AAO17430 Canine en
16	959	77.4	184	3	AAU70265 Canine an
17	940	75.9	684	2	AAU25114 Mouse alp
18	869	70.1	195	3	AAW90874 Human HNW
19	868	70.1	193	3	AAW90877 Human HNW
20	850	68.6	184	5	ABG31794 Murine en
21	843	68.0	191	3	ABAB28398 Murine en
22	843	68.0	191	5	AAU77950 Amino aci
23	838.5	67.7	271	3	AAU08407 A human a
24	836	67.5	215	3	AAU08407 Amino aci
25	835	67.4	181	4	AAU00898 Human End

26	835	67.4	182	3	AAU59622 Human end
27	835	67.4	182	3	AAU94323 Human end
28	835	67.4	182	3	AAU28399 Human end
29	835	67.4	182	4	AAU00897 Human End
30	835	67.4	182	5	AAU77951 Amino aci
31	835	67.4	183	2	AAU02113 SEQ ID 76
32	835	67.4	183	2	AAU08693 Human end
33	835	67.4	183	3	AAU70252 Human ang
34	835	67.4	183	3	AAU90771 Human ang
35	835	67.4	183	3	AAU16451 Human end
36	835	67.4	183	3	AAU30493 Amino aci
37	835	67.4	183	4	AAU49379 Human end
38	835	67.4	183	4	AAU00896 Human end
39	835	67.4	183	5	AAU79901 Human end
40	835	67.4	183	5	AAU49503 Human end
41	835	67.4	183	5	AAU48895 Human end
42	835	67.4	183	5	AAU97132 Human end
43	835	67.4	183	6	AAU79753 Human end
44	835	67.4	513	5	ABG73586 Human End
45	834	67.3	180	4	AAU00899 Human End

ALIGNMENTS

RESULT 1

AAO17429
ID AAO17429 standard; protein; 230 AA.

XX AAO17429;

AC AAO17429;

DT 19-JUL-2002 (first entry)

DE Canine pro-endostatin.

XX

KW

KW

KW

KW

KW

KW

KW

KW

KW

XX

OS

XX

FN

XX

PD

XX

XX

PF

XX

XX

FR

XX

XX

FA

XX

PI

XX

DR

DR

XX

PT

PT

XX

PS

XX

CC

CC

CC

Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
plaque neovascularisation; telangiectasia; haemophilic joints;
angiofibroma; wound granulation; coronary collateral;
ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
cystostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
gynaecological.

Canis familiaris.

EPI191036-A2.

27-MAR-2002.

24-AUG-2001; 2001EP-00307224.

25-AUG-2000; 2000US-0227924P.

(PFIZ) PFIZER PROD INC.

Sheppard MG, Tong X;

WPI; 2002-354068/39.

N-PSDB; AAL46062.

An isolated nucleic acid molecule for the treatment of angiogenesis-related disorder, such as cancers or diabetic retinopathy, encodes an endostatin protein.

Claim 14; Fig 3; 56pp; English.

The present invention provides the protein and coding sequences of canine pro-endostatin and endostatin. The sequences can be used in the treatment and diagnosis of angiogenesis related disorders, including cancer, rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,

CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
 CC rubecsis, Osler-Weber Syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiodiroma,
 CC wound granulation" coronary collaterals, cerebral collaterals,
 CC arteriovenous malformations, ischaemic limb angiogenesis, diabetic
 CC neovascularisation, and fractures. The present sequence is the canine pro
 CC -endostatin protein sequence
 XX
 SQ Sequence 230 AA;

Query Match 100.0%; Score 1239; DB 5; Length 230;
 Best Local Similarity 100.0%; Pred. No. 3.2e-131;
 Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PWRADDILAGPRLLDPQYPGAPHGSGVYHFPAPRTGCVPTHTTHQDFQLVHLVA 60
 DB 1 PWRADDILAGPRLLDPQYPGAPHGSGVYHFPAPRTGCVPTHTTHQDFQLVHLVA 60
 QY 61 LNSPQPGMGIRGADFCQFQARAAAGLAGTFRFLSSRLQDLYSIVRRADRTGVVNL 120
 DB 61 LNSPQPGMGIRGADFCQFQARAAAGLAGTFRFLSSRLQDLYSIVRRADRTGVVNL 120
 QY 121 RDEVLPFSWEALFSGSEGQKPGQARIFSPDGRDVLQHPAMPKSVYHSGDFSGRLTDSY 180
 DB 121 RDEVLPFSWEALFSGSEGQKPGQARIFSPDGRDVLQHPAMPKSVYHSGDFSGRLTDSY 180
 QY 181 CETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVMTSFK 230
 DB 181 CETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVMTSFK 230

RESULT 2
 AAU76689
 ID AAU76689 standard; protein; 275 AA.
 AC AAU76689;
 XX
 DT 21-MAY-2002 (first entry)
 DE
 DE Synthetic plasmid pEnd-HR#1 FPD fusion protein sequence.
 KW Mouse; Ig signal peptide; mlgSP; functional protein domain; FPD;
 KW primary translational product; PTP; DNA construct; regulatory DNA;
 KW DNA targeting segment; regulatory factor; single regulatory unit;
 KW monoclonal antibody; recombination-derived alteration; blood product;
 KW human; COL18A1; mutant; mutein; fusion protein.
 XX

Mus sp.
 OS Homo sapiens.
 OS Synthetic.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /label= Signal_peptide
 FT /note= "Mouse Ig signal peptide (mIgSP)"
 FT Region 1..19
 FT /note= "Encoded by mouse Ig signal peptide (mIgSP) exon"
 FT Protein 20..275
 FT /label= Mature_human_COL18A1_protein
 FT /note= "Contains exons 38-41"

WO200210372-A1.
 XX
 XX
 XX 07-FEB-2002.
 XX
 XX 01-AUG-2001; 2001WO-GB003455.
 XX
 XX 01-AUG-2000; 2000GB-00018876.
 XX
 XX (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
 PA (CHAP//) CHAPMAN P W.
 XX

PI Chapman PW, De Luca G, Falcicola L;
 XX WPI; 2002-195963/25.
 DR N-PSDB; ABK09577.
 XX

PT Producing functional protein domain by growing host cell transfected with
 PT DNA construct having regulatory DNA and DNA targeting segment, and
 PT optionally culturing homologically recombinant cell and collecting
 PT protein.
 XX

XX Example; Fig 8; 116pp; English.

XX The present invention relates to a new method of producing a protein,
 CC such as functional protein domain, that is either C- or N-terminus of the
 CC primary translational product (PTP) of a gene, where the protein has
 CC biological activity which is distinct from PTP. The method of the
 CC invention involves growing a host cell transfected with a DNA construct
 CC comprising a regulatory DNA and a DNA targeting segment. This method is
 CC useful for producing a functional protein domain of proteins such as
 CC regulatory factors, blood products and monoclonal antibodies. The method
 CC described in the invention allows controlled and precise modification of
 CC the host cell genome in order to produce functional protein domain (FPD).
 CC The amount of exogenous sequence to be integrated in the host cell genome
 CC is very limited since, as coding sequence, the original coding sequence
 CC present in the host cell genome itself is used. Use of the host cell
 CC sequence encoding FPD also provides the advantages of both eliminating
 CC any recombination-derived alteration of such coding sequence, and also
 CC making use of the same post-transcriptional (e.g., splicing) and/or post-
 CC translational (e.g. glycosylation, phosphorylation) processes that are
 CC actually applied in vivo for the maturation of FPD. The use of a single
 CC regulatory unit eliminates the necessity of manipulating the
 CC complementary DNA coding for the PTP to isolate the segment coding for
 CC the FPD, and adapt it to the expression vector. The present amino acid
 CC sequence represents the plasmid pEnd-HR#1 FPD fusion protein of the
 CC invention. This fusion protein contains the mouse Ig signal peptide
 CC (mIgSP) sequence fused to exons 38-41 of the human COL18A1 sequence
 XX

SQ Sequence 275 AA;

Query Match 82.2%; Score 1019; DB 5; Length 275;
 Best Local Similarity 82.4%; Pred. No. 3.1e-106;
 Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;
 QY 1 PWRADDILAGPRLLDPQYPGAPHGSGVYHFPAPRTGCVPTHTTHQDFQLVHLVA 60
 DB 51 PWRADDILAGPRLLDPQYPGAPHGSGVYHFPAPRTGCVPTHTTHQDFQLVHLVA 106
 QY 61 LNSPQPGMGIRGADFCQFQARAAAGLAGTFRFLSSRLQDLYSIVRRADRTGVVNL 120
 DB 107 LNSPQPGMGIRGADFCQFQARAAAGLAGTFRFLSSRLQDLYSIVRRADRTGVVNL 166
 QY 121 RDEVLPFSWEALFSGSEGQKPGQARIFSPDGRDVLQHPAMPKSVYHSGDFSGRLTDSY 180
 DB 167 KDELFPFSWEALFSGSEGQKPGQARIFSPDGRDVLQHPAMPKSVYHSGDFSGRLTDSY 226
 QY 181 CETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVMTS 227
 DB 227 CETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVMTA 273

RESULT 3
 AAU76688
 ID AAU76688 standard; protein; 310 AA.
 XX
 AC AAU76688;
 XX
 DT 21-MAY-2002 (first entry)
 XX

DE Human collagen XVIII 1alpha NC1 domain protein sequence.

XX Human; collagen XVIII 1alpha NC1 domain; functional protein domain; FPD;
 KW primary translational product; PTP; DNA construct; regulatory DNA;
 KW DNA targeting segment; regulatory factor; single regulatory unit;
 KW

KW monoclonal antibody; recombination-derived alteration; blood product.
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FH Domain 1..54 /label= Multimerisation_domain
FT Region 1..43 /note= "Encoded by exon 36"
FT Region 44..54 /note= "Encoded by exon 37"
FT Region 55..136 /label= "Hinge region"
FT Region /note= "Encoded by exon 38"
FT Domain 137..310 /label= Endostatin core domain
FT /note= "Autonomous folding unit"
FT Region 137..201 /note= "Encoded by exon 39"
FT Region 202..241 /note= "Encoded by exon 40"
FT Region 242..310 /note= "Encoded by exon 41"
XX
XX WO200210372-A1.
XX
XX
XX 07-FEB-2002.
XX
XX 01-AUG-2001; 2001WO-GB003455.
XX
XX 01-AUG-2000; 2000GB-00018876.
XX
XX (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
XX (CHAP/) CHAPMAN P W.
XX
XX Chapman PW, De Luca G, Falcicola L;
XX
XX WPI; 2002-195963/25.
XX
XX Producing functional protein domain by growing host cell transfected with
XX DNA construct having regulatory DNA and DNA targeting segment, and
XX optionally culturing homologously recombinant cell and collecting
XX protein.
XX
XX Example; Fig 3; 116pp; English.
XX
XX The present invention relates to a new method of producing a protein,
XX such as functional protein domain, that is either C- or N-terminus of the
XX primary translational product (PTP) of a gene, where the protein has
XX biological activity which is distinct from PTP. The method of the
XX invention involves growing a host cell transfected with a DNA construct
XX comprising a regulatory DNA and a DNA targeting segment. This method is
XX useful for producing a functional protein domain of proteins such as
XX regulatory factors, blood products and monoclonal antibodies. The method
XX described in the invention allows controlled and precise modification of
XX the host cell genome in order to produce functional protein domain (FPD).
XX The amount of exogenous sequence to be integrated in the host cell genome
XX is very limited since, as coding sequence, the original coding sequence
XX present in the host cell genome itself is used. Use of the host cell
XX sequence encoding FPD also provides the advantages of both eliminating
XX any recombination-derived alteration of such coding sequence, and also
XX making use of the same post-transcriptional (e.g., splicing) and/or post-
XX translational (e.g., glycosylation, phosphorylation) processes that are
XX actually applied in vivo for the maturation of FPD. The use of a single
XX regulatory unit eliminates the necessity of manipulating the
XX complementary DNA coding for the PTP to isolate the segment coding for
XX the FPD, and adapt it to the expression vector. The present amino acid
XX sequence represents the human collagen XVIII alpha NC1, a functional
XX protein domain used to illustrate the method of the invention
XX
XX Sequence 310 AA;
SQ

Best Local Similarity 82.4%; Pred. No. 3.7e-106;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;
QY 1 PWRADDILAGPRLLDPPQYPGAPHHGSSVYHFQAPRTGPGVHTHTHODFQVLHLVA 60
DB 86 PWRADDILASPPRLPEPQYPGAPHHGSSVYHLRPARPTSP-----AHSRDFQPVHLVA 141
QY 61 LNSPQPGMGEGIRGADPQCFQQAARAAGLAGTFRFLSSRLQDLYSIVRRADRTGVPVNL 120
DB 142 LNSPFGSGMGIRGADPQCFQQAARAAGLAGTFRFLSSRLQDLYSIVRRADRAAIVNL 201
QY 121 RDEVLPFSWEALFSGEGOLKPGARIFSPDGRDVLQHPAPWRKSVYWHGSDPSSGRRLTDSY 180
DB 202 KDELFPFSWEALFSGEGPLKPGARIFSPDGRDVLQHPAPWRKSVYWHGSDPSSGRRLTDSY 261
QY 181 CETWRTAPATQOASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
DB 262 CETWRTAPATQOASSLLAGRLLEQEAASCRHAFVVLCIENSFMFTA 308
RESULT 4
ABP41878
ID ABP41878 standard; protein; 682 AA.
XX
XX AC ABP41878;
XX
XX 22-AUG-2002 (first entry)
XX
XX Human ovarian antigen HEBEK29, SEQ ID NO:3010.
XX
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
XX ovarian cancer; breast cancer; tumour; reproductive system disorder;
XX infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
XX PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
XX inflammatory condition; immune disorder; blood disorder;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disorder; urinary system disorder; drug screening;
XX gene therapy; chromosome mapping; forensic analysis;
XX antibody preparation; cytostatic; immunomodulatory; neuroprotective;
XX antiinflammatory; gynaecological; reproductive; chromosome 21q22.3.
XX
XX Homo sapiens.
XX
XX WO200200677-A1.
XX
XX 03-JAN-2002.
XX
XX 07-JUN-2001; 2001WO-US018569.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX
XX WPI; 2002-147878/19.
XX
XX N-PSDB; ABQ54955.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
XX useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
XX cancer), immune disorders, cardiovascular disorders and neurological
XX diseases.
XX
XX Claim 11; SEQ ID NO 3010; 2922pp; English.
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
XX ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
XX encompasses polypeptides 90% identical and polynucleotides 98% identical
XX to the sequences of the invention. The invention additionally relates to
XX recombinant vectors and host cells comprising human ovarian antigen
XX polynucleotides, antibodies against human ovarian antigens, and the use
XX of ovarian antigen polynucleotides and polypeptides in diagnosing,
XX treating, prognosing or preventing various ovary and/or breast-related

disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia, cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders). Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents a human ovarian antigen of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 82.2%; Score 1019; DB 5; Length 682;
Best Local Similarity 82.4%; Pred. No. 1.1e-105;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;
SQ Sequence 682 AA;
QY 1 PWRADIIAGPRLDDPQYPGAPHGSGYVHFQAPRTGPGVTHTHQDFQVLHLVA 60
DB 458 PWRADIIAGPRLDDPQYPGAPHGSGYVHFQAPRTGPGVTHTHQDFQVLHLVA 513
QY 61 LNSPQPGMGIRGADFCFQQAARAGLAGTTFRAFLSSRLQDLYSIVRRADRTGVPVNL 120
DB 514 LNSPLSGMGIRGADFCFQQAARAGLAGTTFRAFLSSRLQDLYSIVRRADRAVPVNL 573
QY 121 RDEVLPFSWEALFSGSEGQKPGARIFSGDGRVLOHPAMPKSVVHSGDPPSGRRITDSY 180
DB 574 KDELFPFSWEALFSGSEGQKPGARIFSGDGRVLOHPAMPKSVVHSGDPPSGRRITDSY 633
QY 181 CETWTEAPATGQASSILGRLGQSAASCHAYIVLCIENSFMTA 227
DB 634 CETWTEAPATGQASSILGRLGQSAASCHAYIVLCIENSFMTA 680

RESULT 5
AAW92296
ID AAW92296 standard; peptide; 1301 AA.
XX AAW92296;
AC AAW92296;
XX 28-APR-1999 (first entry)
DT Human alpha-1 (XVIII) collagen chain common sequence HUI8(common)36.
XX Human; type XVIII collagen; liver disease; cirrhosis; detection;
KW hepatocellular carcinoma; diagnosis.
XX Homo sapiens.
OS WO9856399-A1.
XX 17-DEC-1998.
XX 12-JUN-1998; 98WO-US012327.
XX 12-JUN-1997; 97US-0049369P.
XX (FIBR-) FIBROGEN INC.
PA (FIB-) ACAD FINLAND.
XX (INEM) INST NAT SANTE & RECH MEDICALE.
XX Pihlajaniemi T, Rehn M, Clement B;

WPI; 1999-070292/06.
Diagnosis and monitoring of liver disease by measuring collagen type XVIII levels - with elevated levels indicative of disease, especially cirrhosis or hepatocellular carcinoma.
Example 6; Fig 8; 56pp; English.
A method has been developed for the detecting liver disease. The method comprises: (a) reacting a patient sample with antibodies (Ab) specific for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen complex (C) formed as indicator of the amount of Coll18 present; (c) similar analysis of a non-diseased control; and (d) comparing the amounts of Coll18 in the two samples to detect presence or progression of disease. Elevated levels of Coll18 are: (i) indicative of disease, specifically cirrhosis; and (ii) predictive of the prognosis of disease, specifically hepatocellular carcinoma (there is a relationship between Coll18 mRNA levels and tumour size and necrosis, and survival times are significantly higher in patients with higher Coll18 levels). The method provides non-invasive, early and accurate diagnosis of liver disease. The present sequence represents the sequence common to human alpha-1 (XVIII) collagen chain from the present invention

Query Match 82.2%; Score 1019; DB 2; Length 1301;
Best Local Similarity 82.4%; Pred. No. 2.9e-105;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;
SQ Sequence 1301 AA;
QY 1 PWRADIIAGPRLDDPQYPGAPHGSGYVHFQAPRTGPGVTHTHQDFQVLHLVA 60
DB 1077 PWRADIIAGPRLDDPQYPGAPHGSGYVHFQAPRTGPGVTHTHQDFQVLHLVA 1132
QY 61 LNSPQPGMGIRGADFCFQQAARAGLAGTTFRAFLSSRLQDLYSIVRRADRTGVPVNL 120
DB 1133 LNSPLSGMGIRGADFCFQQAARAGLAGTTFRAFLSSRLQDLYSIVRRADRAVPVNL 1192
QY 121 RDEVLPFSWEALFSGSEGQKPGARIFSGDGRVLOHPAMPKSVVHSGDPPSGRRITDSY 180
DB 1193 KDELFPFSWEALFSGSEGQKPGARIFSGDGRVLOHPAMPKSVVHSGDPPSGRRITDSY 1252
QY 181 CETWTEAPATGQASSILGRLGQSAASCHAYIVLCIENSFMTA 227
DB 1253 CETWTEAPATGQASSILGRLGQSAASCHAYIVLCIENSFMTA 1299

RESULT 6
AAW08694
ID AAW08694 standard; protein; 1336 AA.
XX AAW08694;
AC AAW08694;
XX 10-AUG-1999 (first entry)
DT Human collagen 18 protein.
XX Human collagen 18 protein.
DE Plasminogen; human; angiotensin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
XX tumour growth; solid tumour; diabetic retinopathy; retina; collagen 18.
XX Homo sapiens.
OS WO9926480-A1.
XX 03-JUN-1999.
XX 20-NOV-1998; 98WO-US024950.
XX 20-NOV-1997; 97US-00975424.
XX (GENE-) GENETIX PHARM INC.
PA (WASI) MASSACHUSETTS INST TECHNOLOGY.

XX WPI: 2002-583596/62.
DR N-PSDB; AEN85301.
XX
PT Novel composition for gene therapy against rheumatoid arthritis,
XX comprising a DNA encoding anti-angiogenic protein or its parts.
XX
PS Claim 4; Page 70-78; 84pp; English.
XX
XX The present invention relates to a composition for gene therapy,
CC comprising a DNA encoding an anti-angiogenic protein, which shows
CC therapeutic effects on rheumatoid arthritis. The composition is useful
CC for treating rheumatoid arthritis and the gene therapy is effective,
CC lasting for 14 days. The present sequence is the protein sequence for
CC human collagen XVIII. Endostatin, which consists of the C-terminal 183
CC residues of collagen XVIII, was used as an anti-angiogenic protein
XX
XX Sequence 1516 AA;
XX
Query Match 82.2%; Score 1019; DB 5; Length 1516;
Best Local Similarity 82.4%; Pred. No. 3.6e-105;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;
XX
QY 1 PWRADDILAGPRLDPPQYPGAPHHGSHYHFPQAPRTGGPVHTHTHQQDFQLVHLVA 60
DB 1292 PWRADDILAGPRLDPPQYPGAPHHGSHYHFPQAPRTGGPVHTHTHQQDFQLVHLVA 1347
QY 61 LNSPQGMGIRGADPQCQQAARAGLAGTFRFLSSRLQDLYSIVRRADRAAVPIVNL 120
DB 1348 LNSPQGMGIRGADPQCQQAARAGLAGTFRFLSSRLQDLYSIVRRADRAAVPIVNL 1407
QY 121 RDEVLFPSWEALFSGSGQLKPGARIFSGDGRDLQHPAPRKSVHSGDSPGRLTDSY 180
DB 1408 KDELLFPSWEALFSGSGQLKPGARIFSGDGRDLQHPAPRKSVHSGDSPGRLTDSY 1467
QY 181 CETWTEAPATGQASSLLAGRLLEQEAASCSHAFVVLCIENSVMTS 227
DB 1468 CETWTEAPATGQASSLLAGRLLEQEAASCSHAFVVLCIENSVMTS 1514
RESULT 9
ABP68617
ID ABP68617 standard; protein; 1516 AA.
XX
AC ABP68617;
XX
DT 14-JAN-2003 (first entry)
XX
DE Human pancreatic cancer expressed protein SEQ ID NO 166.
XX
KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
XX cytosolic; tumour.
XX
OS Homo sapiens.
XX
FN WO200260317-A2.
XX
PD 08-AUG-2002.
XX
PF 30-JAN-2002; 2002WO-US002781.
XX
PR 30-JAN-2001; 2001US-0265305P.
XX
PR 31-JAN-2001; 2001US-0265682P.
XX
PR 09-FEB-2001; 2001US-0267568P.
XX
PR 21-MAR-2001; 2001US-0278651P.
XX
PR 28-APR-2001; 2001US-0287112P.
XX
PR 16-MAY-2001; 2001US-0291631P.
XX
PR 12-JUL-2001; 2001US-0305484P.
XX
PR 20-AUG-2001; 2001US-0313999P.
XX
PR 27-NOV-2001; 2001US-0333626P.
XX
PA (CORI-) CORIXA CORP.
XX

PI Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;
XX WPI: 2002-627435/67.
DR N-PSDB; ABV94763.
XX
PT New isolated polynucleotide and pancreatic tumor polypeptides, useful for
XX diagnosing, preventing and/or treating cancer, particularly pancreatic
XX cancer.
XX
PS Claim 2; SEQ ID NO 166; 300pp + Sequence Listing; English.
XX
XX The invention relates to an isolated polynucleotide (I) comprising: (a)
CC any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)
CC complements of (a); (c) sequences that hybridize to (a) under moderately
CC stringent conditions; (d) sequences that hybridize to (a) under moderately
CC stringent conditions; (e) sequences having at least 75% or 90% identity
CC to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-
CC ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer
CC in a patient and compositions comprising polypeptides, polynucleotides,
CC antibodies, fusion proteins, T cell populations and antigen presenting
CC cells expressing the polypeptide are useful in treating pancreatic cancer
CC and stimulating an immune response. The polynucleotides can be used as
CC probes or primers for nucleic acid hybridization, in the design and
CC preparation of ribozyme molecules for inhibiting expression of the tumour
CC polypeptides and proteins in the tumour cells, in vaccines and for gene
CC therapy. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1516 AA;
XX
Query Match 82.2%; Score 1019; DB 5; Length 1516;
Best Local Similarity 82.4%; Pred. No. 3.6e-105;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;
XX
QY 1 PWRADDILAGPRLDPPQYPGAPHHGSHYHFPQAPRTGGPVHTHTHQQDFQLVHLVA 60
DB 1292 PWRADDILAGPRLDPPQYPGAPHHGSHYHFPQAPRTGGPVHTHTHQQDFQLVHLVA 1347
QY 61 LNSPQGMGIRGADPQCQQAARAGLAGTFRFLSSRLQDLYSIVRRADRAAVPIVNL 120
DB 1348 LNSPQGMGIRGADPQCQQAARAGLAGTFRFLSSRLQDLYSIVRRADRAAVPIVNL 1407
QY 121 RDEVLFPSWEALFSGSGQLKPGARIFSGDGRDLQHPAPRKSVHSGDSPGRLTDSY 180
DB 1408 KDELLFPSWEALFSGSGQLKPGARIFSGDGRDLQHPAPRKSVHSGDSPGRLTDSY 1467
QY 181 CETWTEAPATGQASSLLAGRLLEQEAASCSHAFVVLCIENSVMTS 227
DB 1468 CETWTEAPATGQASSLLAGRLLEQEAASCSHAFVVLCIENSVMTS 1514
RESULT 10
AAW26327
ID AAW26327 standard; protein; 684 AA.
XX
AC AAW26327;
XX
DT 19-NOV-1997 (first entry)
XX
DE Human alpha-1 collagen (XVIII).
XX
XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..6 /label= GXYGX'Y' motif
FT /note= "Claim 1"
FT Peptide 7..12 /label= GXYGX'Y' motif
FT /note= "Claim 1"

FT	Peptide	13..18	/label= GXYGX'Y'_motif	FT	Peptide	193..198	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	19..24	/label= GXYGX'Y'_motif	FT	Peptide	215..220	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	25..30	/label= GXYGX'Y'_motif	FT	Peptide	221..226	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	31..36	/label= GXYGX'Y'_motif	FT	Peptide	227..232	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	37..42	/label= GXYGX'Y'_motif	FT	Peptide	233..238	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	48..53	/label= GXYGX'Y'_motif	FT	Peptide	239..244	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	54..59	/label= GXYGX'Y'_motif	FT	Peptide	257..262	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	74..79	/label= GXYGX'Y'_motif	FT	Peptide	263..268	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	80..85	/label= GXYGX'Y'_motif	FT	Peptide	269..274	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	86..91	/label= GXYGX'Y'_motif	FT	Peptide	275..280	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	92..97	/label= GXYGX'Y'_motif	FT	Peptide	286..291	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	98..103	/label= GXYGX'Y'_motif	FT	Peptide	292..297	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	104..109	/label= GXYGX'Y'_motif	FT	Peptide	298..303	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	110..115	/label= GXYGX'Y'_motif	FT	Peptide	309..314	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	121..126	/label= GXYGX'Y'_motif	FT	Peptide	315..320	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	129..134	/label= GXYGX'Y'_motif	FT	Peptide	322..328	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	135..140	/label= GXYGX'Y'_motif	FT	Peptide	329..334	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	141..146	/label= GXYGX'Y'_motif	FT	Peptide	335..340	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	147..152	/label= GXYGX'Y'_motif	FT	Peptide	354..359	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	153..158	/label= GXYGX'Y'_motif	FT	Peptide	360..365	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	159..164	/label= GXYGX'Y'_motif	FT	Peptide	366..372	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	165..170	/label= GXYGX'Y'_motif	FT	Peptide	523..528	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	171..176	/label= GXYGX'Y'_motif	FT	Peptide	542..547	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/note= "Claim 1"	
FT	Peptide	181..186	/label= GXYGX'Y'_motif	FT	Peptide	590..595	/label= GXYGX'Y'_motif
FT		/note= "Claim 1"		FT		/label= GXYGX'Y'_motif	
FT	Peptide	187..192					

/note= "Claim 1"

FT XX US5643783-A.
 XX PN 01-JUL-1997.
 XX PD 01-DEC-1993; 93US-00159784.
 XX PF 01-DEC-1993; 93US-00159784.
 XX PR (HARD) HARVARD COLLEGE.
 XX PA Olsen BR, Oh SP;
 XX PI WPI; 1997-350247/32.
 XX DR N-PSDB; AAT84484.
 XX XX
 XX PT Nucleic acid encoding human alpha-1 collagen - for production of
 XX PT recombinant alpha-1 collagen, for use in the treatment of cartilage
 XX PT degeneration.
 XX PS Claim 1; Col 23-30; 35pp; English.
 XX CC Novel human type alpha-1 (XVIII) collagen is characterized by 10 triple
 CC helical domains containing the GYXG'Y' motif (where X, Y, X' and Y'
 CC represent any amino acid), the helical domains being separated and
 CC flanked by non-triple helical regions which may provide flexibility.
 CC Alpha-1 collagen is expressed in multiple tissues, especially liver, lung
 CC and kidney. A claimed plasmid comprising alpha-1 collagen nucleic acid
 CC (see AAT84484) and an expression control sequence can be used to express
 CC recombinant collagen in prokaryotic or eukaryotic (especially mammalian)
 CC host cells. The alpha-1 collagen may be used to treat a patient suffering
 CC from a disease associated with cartilage degradation, and for
 CC supplementing collagen. It can also be used as a connective tissue filler
 CC (e.g. for plastic surgery), can be interposed between a dermal equivalent

Query Match 81.0%; Score 1003; DB 2; Length 684;
 Best Local Similarity 81.5%; Pred. No. 7.4e-104;
 Matches 185; Conservative 15; Mismatches 23; Indels 4; Gaps 1;
 QY 1 PWRADILAGPRLDPOYPGAPHGYSYVHFQPARPTGGVHTHTHODFQVLHLVA 60
 DB 460 PWRADILASPGLPEPQYFGGPHSSYVCGPARPTSP---AHSRDFQVHLVA 515
 QY 61 LNSPQPGMRGIRGADFQCFQQAARAAGLAGTFRAPLSSRLQDLYSIVRRADRTGVVNL 120
 DB 516 LNSPLSGMGRGIRGADFQCFQQAARAAGLAGTFRAPLSSRLQDLYSIVRRADRAAPVNL 575
 QY 121 RDEVLPFSWEALFSGSEGLKPGARIFSDGRDVLQHPAPWPKSVHSGSDPSGRRLTDSY 180
 DB 576 KDELLFPSWEALFSGSEGLKPGARIFSDGRDVLQHPAPWPKSVHSGSDPSGRRLTESY 635
 QY 181 CETWTEAPAAATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVMTS 227
 DB 636 CETWTEAPSATGQASSLLGRLGQSAASCHHAYIVLICIENSFMTA 682

RESULT 11
 AAO17357
 ID AAO17357 standard; protein; 684 AA.
 AC AAO17357;
 XX 19-JUL-2002 (first entry)
 DT Human collagen type XVIII alpha 1.
 DE Human; endometriosis; DNA chip; fibronectin; p27; reticulocalbin;
 KW aldehyde dehydrogenase 6; gravin; phospholipase C epsilon; elastin;
 KW insulin-like growth factor binding protein-2; alpha-2 type IV collagen;
 KW transmembrane receptor PK7; collagen type XVIII alpha 1;
 KW platelet derived growth factor receptor alpha; laminin M chain;
 KW subtilisin like protein PACE4; nidogen.

XX OS Homo sapiens.
 XX PN EP1191107-A2.
 XX XX 27-MAR-2002.
 XX PF 21-AUG-2001; 2001EP-00250300.
 XX XX 25-SEP-2000; 2000DE-01048633.
 XX PR (SCHD) SCHERING AG.
 XX PA Hess-Stump H, Haendler B, Kraetzschmar J, Kreft B, Winterhager E;
 XX PI Regidor P, Scotti S;
 XX XX WPI; 2002-317413/36.
 XX DR In vitro diagnosis and monitoring of endometriosis, comprises detecting
 XX PT reduced expression of specific gene products, e.g. from the fibronectin
 XX PT gene.
 XX PS Claim 1; Page 12-13; 21pp; German.
 XX CC The present invention relates to a method for the in vitro diagnosis of
 CC endometriosis by determining the amount of gene product from at least one
 CC specific gene in a patient sample and comparing this with the amount of
 CC gene product in a control sample. A reduced level is indicative of
 CC endometriosis. The gene products may be fibronectin, p27, reticulocalbin,
 CC aldehyde dehydrogenase 6, gravin, phospholipase C epsilon, elastin,
 CC insulin-like growth factor binding protein-2, alpha-2 type IV collagen,
 CC transmembrane receptor PK7, collagen type XVIII alpha 1, platelet
 CC derived growth factor receptor alpha, laminin M chain, subtilisin like
 CC protein PACE4 or nidogen. The method is useful for initial diagnosis of
 CC endometriosis, and also for monitoring progress and treatment of the
 CC disease. The present sequence is human collagen type XVIII alpha 1

Query Match 81.0%; Score 1003; DB 5; Length 684;
 Best Local Similarity 81.5%; Pred. No. 7.4e-104;
 Matches 185; Conservative 15; Mismatches 23; Indels 4; Gaps 1;
 QY 1 PWRADILAGPRLDPOYPGAPHGYSYVHFQPARPTGGVHTHTHODFQVLHLVA 60
 DB 460 PWRADILASPGLPEPQYFGGPHSSYVCGPARPTSP---AHSRDFQVHLVA 515
 QY 61 LNSPQPGMRGIRGADFQCFQQAARAAGLAGTFRAPLSSRLQDLYSIVRRADRTGVVNL 120
 DB 516 LNSPLSGMGRGIRGADFQCFQQAARAAGLAGTFRAPLSSRLQDLYSIVRRADRAAPVNL 575
 QY 121 RDEVLPFSWEALFSGSEGLKPGARIFSDGRDVLQHPAPWPKSVHSGSDPSGRRLTDSY 180
 DB 576 KDELLFPSWEALFSGSEGLKPGARIFSDGRDVLQHPAPWPKSVHSGSDPSGRRLTESY 635
 QY 181 CETWTEAPAAATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVMTS 227
 DB 636 CETWTEAPSATGQASSLLGRLGQSAASCHHAYIVLICIENSFMTA 682

RESULT 12
 AAY25113
 ID AAY25113 standard; protein; 684 AA.
 AC AAY25113;
 XX 25-AUG-1999 (first entry)
 DT Human alpha1 (XVIII) collagen protein.
 DE Human; endometriosis; DNA chip; fibronectin; p27; reticulocalbin;
 KW aldehyde dehydrogenase 6; gravin; phospholipase C epsilon; elastin;
 KW insulin-like growth factor binding protein-2; alpha-2 type IV collagen;
 KW transmembrane receptor PK7; collagen type XVIII alpha 1;
 KW platelet derived growth factor receptor alpha; laminin M chain;
 KW subtilisin like protein PACE4; nidogen.

FT Peptide 548. .553 /label= GXYGX'Y' _motif
FT 580. 585 /label= GXYGX'Y' _motif
FT 586. 591 /label= GXYGX'Y' _motif
FT 592. 597 /label= GXYGX'Y' _motif
FT 598. 603 /label= GXYGX'Y' _motif
FT 604. 609 /label= GXYGX'Y' _motif
FT 610. 615 /label= GXYGX'Y' _motif
FT 616. 621 /label= GXYGX'Y' _motif
FT 622. 627 /label= GXYGX'Y' _motif
FT 628. 633 /label= GXYGX'Y' _motif
FT 634. 639 /label= GXYGX'Y' _motif
FT 640. 665 /label= GXYGX'Y' _motif
FT 657. 662 /label= GXYGX'Y' _motif
FT 677. 682 /label= GXYGX'Y' _motif
FT 683. 688 /label= GXYGX'Y' _motif
FT 689. 694 /label= GXYGX'Y' _motif
FT 695. 700 /label= GXYGX'Y' _motif
FT 707. 712 /label= GXYGX'Y' _motif
FT 713. 718 /label= GXYGX'Y' _motif
FT 735. 740 /label= GXYGX'Y' _motif
FT 741. 746 /label= GXYGX'Y' _motif
FT 747. 752 /label= GXYGX'Y' _motif
FT 759. 764 /label= GXYGX'Y' _motif
FT 765. 770 /label= GXYGX'Y' _motif
FT 771. 776 /label= GXYGX'Y' _motif
FT 787. 792 /label= GXYGX'Y' _motif
FT 793. 798 /label= GXYGX'Y' _motif
FT 799. 804 /label= GXYGX'Y' _motif
FT 815. 820 /label= GXYGX'Y' _motif
FT 821. 826 /label= GXYGX'Y' _motif
FT 827. 832 /label= GXYGX'Y' _motif
FT 833. 838 /label= GXYGX'Y' _motif
FT 839. 844 /label= GXYGX'Y' _motif
FT 845. 850 /label= GXYGX'Y' _motif
FT 863. 868 /label= GXYGX'Y' _motif
FT 869. 874 /label= GXYGX'Y' _motif
FT 875. 880 /label= GXYGX'Y' _motif

FT Peptide /label= GXYGX'Y' _motif
FT 891. 896 /label= GXYGX'Y' _motif
FT 897. 902 /label= GXYGX'Y' _motif
FT 903. 908 /label= GXYGX'Y' _motif
FT 911. 916 /label= GXYGX'Y' _motif
FT 917. 922 /label= GXYGX'Y' _motif
FT 928. 933 /label= GXYGX'Y' _motif
FT 934. 939 /label= GXYGX'Y' _motif
FT 956. 961 /label= GXYGX'Y' _motif
FT 962. 967 /label= GXYGX'Y' _motif
FT 968. 973 /label= GXYGX'Y' _motif
FT 1126. 1131 /label= GXYGX'Y' _motif
FT 1145. 1150 /label= GXYGX'Y' _motif
FT 1193. 1198 /label= GXYGX'Y' _motif

XX US5643783-A.

XX 01-JUL-1997.

XX 01-DEC-1993; 93US-00159784.

XX 01-DEC-1993; 93US-00159784.

XX (HARD) HARVARD COLLEGE.

XX Olsen BR, Oh SP;

XX WPI; 1997-350247/32.

XX N-PSDB; AAT84485.

XX Nucleic acid encoding human alpha-1 collagen - for production of recombinant alpha-1 collagen, for use in the treatment of cartilage degeneration.

PS Disclosure; Fig 2; 35pp; English.

Query Match 79.1%; Score 980.5; DB 2; Length 1288;
Best Local Similarity 79.7%; Pred No. 6.3e-101;
Matches 184; Conservative 19; Mismatches 23; Indels 5; Gaps 2;

QY 1 PWRADDILAGPRLDPPQYPGAP-HGGSYVHFQPARPTGGPVHTHTHDSFQVLHLV 59
DB 1062 PWRADDILANPRLPDRQYPGVPHHSSVHLPPARPET---LSLAHTHDSFQVLHLV 1117
QY 60 ALNSPQFGMRGIRGADFOCFQARAAGLAGTFRALSSRLQDLYSIIVRRADRTGVPVN 119
DB 1118 ALNPLSGMRGIRGADFOCFQARAAGLAGTFRALSSRLQDLYSIIVRRADRTGVPVN 1177
QY 120 LRDEVLFPSSWEALFSGEGQLKPGARIFSDGRDYLQHPAMPRKSVHSGDPSGRRRLTDS 179
DB 1178 LKDEVLSPSWDSLFSGSGQGVQPGARIFSDGRDYLQHPAMPRKSVHSGDPSGRRRLMES 1237
QY 180 YCETWRTTEAPAAATGQASSLLAGRLLEQEAASCHAFVVLICIENSVMTSFSK 230
DB 1238 YCETWRTTEATGQASSLLAGRLLEQEAASCHNSIIVLCIENSVMTSFSK 1288

RESULT 14
AAW92297
ID AAW92297 standard; peptide; 1288 AA.

XX AAW922297;
XX 28-APR-1999 (first entry)
XX Mouse alpha-1 (XVIII) collagen chain common sequence MO18 (common)28.
XX Human; type XVIII collagen; liver disease; cirrhosis; detection;
XX hepatocellular carcinoma; diagnosis.
XX Mus sp.
XX WO9856399-A1.
XX 17-DEC-1998.
XX 12-JUN-1998; 98WO-US012327.
XX 12-JUN-1997; 97US-0049369P.
XX (FIBR-) FIBROGEN INC.
XX (FIFI-) ACAD FINLAND.
XX (INRM) INST NAT SANTE & RECH MEDICALE.
XX Pihlajaniemi T, Rehn M, Clement B;
XX WFI; 1993-070292/06.
XX Diagnosis and monitoring of liver disease by measuring collagen type
XX XVIII levels - with elevated levels indicative of disease, especially
XX cirrhosis or hepatocellular carcinoma.
XX Example 6; Fig 8; 56pp; English.
XX A method has been developed for the detecting liver disease. The method
XX comprises: (a) reacting a patient sample with antibodies (Ab) specific
XX for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen
XX complex (C) formed as indicator of the amount of Coll18 present; (c)
XX similar analysis of a non-diseased control; and (d) comparing the amounts
XX of Coll18 in the two samples to detect presence or progression of disease.
XX Elevated levels of Coll18 are: (i) indicative of disease, specifically
XX cirrhosis; and (ii) predictive of the prognosis of disease, specifically
XX hepatocellular carcinoma (there is a relationship between Coll18 mRNA
XX levels and tumour size and necrosis, and survival times are significantly
XX higher in patients with higher Coll18 levels). The method provides non-
XX invasive, early and accurate diagnosis of liver disease. The present
XX sequence represents the sequence common to mouse alpha-1 (XVIII) collagen
XX chain from the present invention
XX
XX Sequence 1288 AA;
Query Match 78.8%; Score 976; DB 2; Length 1288;
Best Local Similarity 79.7%; Pred. No. 2e-100;
Matches 184; Conservative 20; Mismatches 21; Indels 6; Gaps 3;
QY 1 PWRADDILAGPRLDPQPYGAP-HGGSYVHF-QPARPTGGFVTHTHTHQDFQFVLHL 58
DB 1062 PWRADDILANPRLDPQPYGVPVPHHSSVYHLPPREPRT-LSLARHTQDFQFVLHL 1117
QY 59 VALNSPQPGWGRIGADFCFQQAARAAGLAGTFRFLSSRLQDLYSIVRRADRTGVVW 118
DB 1118 VALNTPUSGWRGIRGADFCFQQAARAAGLAGTFRFLSSRLQDLYSIVRRADRTGVV 1177
QY 119 NLRDEVLPFSWEALFSGSEQLKPGARIFSGDRDVLQHPAMPKRSVWHGSDPSGRRITD 178
DB 1178 NLKDEVLPSSWDSIFSGSQGVQPGARIFSGDRDVLQHPAMPKRSVWHGSDPSGRRILME 1237
QY 179 SYCETWTETPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMFS 229
DB 1238 SYCETWTETTTGQASSLLSGLLEQKAASCHNSYIVLCIENSFMFS 1288

RESULT 15

AAO17430
ID AAO17430 standard; protein; 184 AA.
XX
XX AAO17430;
XX
DT 19-JUL-2002 (first entry)
XX
XX Canine endostatin.
XX
XX Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
XX psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
XX corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
XX rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
XX plaque neovascularisation; telangiectasia; haemophilic joints;
XX angiofibroma; wound granulation; coronary collateral;
XX cerebral collateral; arteriovenous malformation;
XX ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
XX cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
XX gynaecological.
XX
XX Canis familiaris.
XX
XX EP1191036-A2.
XX
XX 27-MAR-2002.
XX
XX 24-AUG-2001; 2001EP-00307224.
XX
XX 25-AUG-2000; 2000US-0227924P.
XX
XX (PFIZ) PFIZER PROD INC.
XX
XX Sheppard MG, Tong X;
XX
XX WFI; 2002-354068/39.
XX
XX N-PSDB; AAL46053.
XX
XX An isolated nucleic acid molecule for the treatment of angiogenesis-
XX related disorder, such as cancers or diabetic retinopathy, encodes an
XX endostatin protein.
XX
XX Claim 14; Fig 5; 56pp; English.
XX
XX The present invention provides the protein and coding sequences of canine
XX pro-endostatin and endostatin. The sequences can be used in the treatment
XX and diagnosis of angiogenesis related disorders, including cancer,
XX rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
XX corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
XX rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
XX neovascularisation, telangiectasia, haemophilic joints, angiobroma,
XX wound granulation, coronary collaterals, cerebral collaterals,
XX arteriovenous malformations, ischaemic limb angiogenesis, diabetic
XX neovascularisation, and fractures. The present sequence is the canine pro
XX endostatin protein sequence
XX
XX Sequence 184 AA;
Query Match 78.0%; Score 966; DB 5; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.7e-100;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 HTHQDFQFVLHLVALNSPQPGWGRIGADFCFQQAARAAGLAGTFRFLSSRLQDLYSI 106
DB 1 HTHQDFQFVLHLVALNSPQPGWGRIGADFCFQQAARAAGLAGTFRFLSSRLQDLYSI 60
QY 107 VRRADRTGVVWNLRDVLPFSWEALFSGSEQLKPGARIFSGDRDVLQHPAMPKRSVW 166
DB 61 VRRADRTGVVWNLRDVLPFSWEALFSGSEQLKPGARIFSGDRDVLQHPAMPKRSVW 120
QY 167 HGSDDPSGRRITSYCETWTETPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 226
DB 121 HGSDDPSGRRITSYCETWTETPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180

QY 227 SFSK 230
|||
Db 181 SFSK 184

Search completed: March 26, 2004, 13:35:56
Job time : 60.2222 secs

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:34:07 ; Search time 18.3333 Seconds
(without alignments)
647.671 Million cell updates/sec

Title: US-09-938-391-2

Perfect score: 1239

Sequence: 1 PWRADDLAGPRLDPQPY.....CRHAFVVLCTNSVMTSFSK 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	843	68.0	191	4	US-09-561-500-13
2	843	68.0	191	4	US-09-561-108-13
3	843	68.0	191	4	US-09-561-526-13
4	843	68.0	191	4	US-09-561-499-13
5	843	68.0	191	4	US-09-998-831-13
6	839	67.7	195	1	US-08-159-784-2
7	835	67.4	182	4	US-08-561-500-14
8	835	67.4	182	4	US-08-561-108-14
9	835	67.4	182	4	US-08-315-589-3
10	835	67.4	182	4	US-08-561-526-14
11	835	67.4	182	4	US-08-561-499-14
12	835	67.4	182	4	US-09-998-831-14
13	835	67.4	183	3	US-08-206-059-2
14	817	65.9	178	4	US-09-315-689-5
15	790	63.8	185	3	US-08-985-526-36
16	513	41.4	191	1	US-08-159-784-3
17	195.5	15.8	124	4	US-08-231-077D-10
18	188	15.2	123	4	US-08-231-077D-11
19	148	11.9	35	3	US-09-046-985-2
20	148	11.9	35	3	US-09-474-743-2
21	101	8.2	22	3	US-09-046-985-7
22	101	8.2	22	3	US-09-474-743-7
23	98.5	7.9	369	4	US-09-252-991A-25533
24	97	7.8	16	3	US-09-385-442-32
25	96	7.7	322	4	US-09-252-991A-27326
26	94.5	7.6	795	4	US-09-252-991A-19085
27	93	7.5	493	4	US-09-252-991A-23421

28 91 7.3 618 4 US-09-252-991A-28358 Sequence 28358, A
29 90.5 7.3 757 3 US-09-413-814-84 Sequence 84, Appl
30 90 7.3 534 4 US-09-252-991A-22537 Sequence 22537, A
31 89.5 7.2 432 4 US-09-252-991A-24690 Sequence 24690, A
32 88.5 7.1 630 4 US-09-252-991A-19702 Sequence 19702, A
33 87.5 7.1 541 4 US-09-252-991A-17206 Sequence 17206, A
34 87.5 7.1 816 2 US-08-267-803B-9 Sequence 9, Appl
35 87.5 7.1 816 3 US-09-041-886-17 Sequence 17, Appl
36 87 7.0 478 4 US-09-252-991A-28496 Sequence 28496, A
37 87 7.0 490 4 US-09-252-991A-21383 Sequence 21383, A
38 87 7.0 512 4 US-09-252-991A-17586 Sequence 17586, A
39 87 7.0 731 4 US-09-252-991A-17180 Sequence 17180, A
40 86.5 7.0 430 4 US-09-252-991A-16912 Sequence 16912, A
41 86.5 7.0 654 4 US-09-307-744A-177 Sequence 177, App
42 86.5 7.0 654 4 US-09-905-125A-177 Sequence 177, App
43 86.5 7.0 654 4 US-09-902-775A-177 Sequence 177, App
44 86 6.9 20 2 US-08-740-168A-1 Sequence 1, Appl
45 86 6.9 20 3 US-09-349-429-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-561-500-13

; Sequence 13, Application US/09561500

; Patent No. 6342219

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002500

; CURRENT APPLICATION NUMBER: US/09/561,500

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 191

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-561-500-13

Query Match 68.0%; Score 843; DB 4; Length 191;
Best Local Similarity 83.0%; Pred. No. 5.6e-87;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

Qy 43 HTHTHTQDFQVLVHLVALNSPPQGMGIRGADFCQFQQAAGLAGTFFAFLLSSRLQD 102
Db 4 HHHHTHTQDFQVLVHLVALNTPLSGMGIRGADFCQFQQAAGLAGTFFAFLLSSRLQD 63
Qy 103 LYSIVRADRTGVVNLRLDEVLPFSEWALFSGEGQLKPGARIPSPDGRVLOHPAMPR 162
Db 64 LYSIVRADRTGVVNLRLDEVLPFSEWALFSGEGQLKPGARIPSPDGRVLOHPAMPR 123
Qy 163 KSVWHSDFSGRLTDSYCEWTWRTPAPATGQASLLAGRLLEQBAASCRHAFVVLCTN 222
Db 124 KSVWHSDFSGRLTDSYCEWTWRTPAPATGQASLLAGRLLEQBAASCRHAFVVLCTN 183
Qy 223 SVMTSFSK 230
Db 184 SPMTSFSK 191

RESULT 2

US-09-561-108-13

; Sequence 13, Application US/09561108

; Patent No. 6342221

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-13

Query Match 68.0%; Score 843; DB 4; Length 191;
Best Local Similarity 83.0%; Pred. No. 5.6e-87;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

QY 43 HTHHTHQQDFQVVLHVALNSPQGMGIRGADFCQFQQAARAGLAGTFFRAFLSSRLQD 102
DB 4 HHHHTHQQDFQVVLHVALNTPLSGGMRGIRGADFCQFQQAARAVGLSGTFFRAFLSSRLQD 63

QY 103 LYSIVRRADRTGVPVNLDEVLPFSWEALFSGSEQLKPGARIFSGDGRDVLQHPAWPR 162
DB 64 LYSIVRRADRTGVPVNLKDEVLPSPWDSLFGSQQLQPGARIFSGDGRDVLQHPAWPR 123

QY 163 KSVWHGSDPSGRRLTDSYCYETWTETAPATGQASSLLAGRLLEQAAACRHAFFVVLCIEN 222
DB 124 KSVWHGSDPSGRRLMESYCYETWTETGATGQASSLLSGLRLEQAAASCHNSYIVLCIEN 183

QY 223 SVMTSFSK 230
DB 184 SFMTSFSK 191

RESULT 3
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Rolf A. Brekken
; APPLICANT: Philip E. Thorpe
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-526-13

Query Match 68.0%; Score 843; DB 4; Length 191;
Best Local Similarity 83.0%; Pred. No. 5.6e-87;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

QY 43 HTHHTHQQDFQVVLHVALNSPQGMGIRGADFCQFQQAARAGLAGTFFRAFLSSRLQD 102
DB 4 HHHHTHQQDFQVVLHVALNTPLSGGMRGIRGADFCQFQQAARAVGLSGTFFRAFLSSRLQD 63

QY 103 LYSIVRRADRTGVPVNLDEVLPFSWEALFSGSEQLKPGARIFSGDGRDVLQHPAWPR 162
DB 64 LYSIVRRADRTGVPVNLKDEVLPSPWDSLFGSQQLQPGARIFSGDGRDVLQHPAWPR 123

QY 163 KSVWHGSDPSGRRLTDSYCYETWTETAPATGQASSLLAGRLLEQAAACRHAFFVVLCIEN 222
DB 124 KSVWHGSDPSGRRLMESYCYETWTETGATGQASSLLSGLRLEQAAASCHNSYIVLCIEN 183

QY 223 SVMTSFSK 230
DB 184 SFMTSFSK 191

RESULT 4
US-09-561-499-13
; Sequence 13, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-499-13

Query Match 68.0%; Score 843; DB 4; Length 191;
Best Local Similarity 83.0%; Pred. No. 5.6e-87;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

QY 43 HTHHTHQQDFQVVLHVALNSPQGMGIRGADFCQFQQAARAGLAGTFFRAFLSSRLQD 102
DB 4 HHHHTHQQDFQVVLHVALNTPLSGGMRGIRGADFCQFQQAARAVGLSGTFFRAFLSSRLQD 63

QY 103 LYSIVRRADRTGVPVNLDEVLPFSWEALFSGSEQLKPGARIFSGDGRDVLQHPAWPR 162
DB 64 LYSIVRRADRTGVPVNLKDEVLPSPWDSLFGSQQLQPGARIFSGDGRDVLQHPAWPR 123

QY 163 KSVWHGSDPSGRRLTDSYCYETWTETAPATGQASSLLAGRLLEQAAACRHAFFVVLCIEN 222
DB 124 KSVWHGSDPSGRRLMESYCYETWTETGATGQASSLLSGLRLEQAAASCHNSYIVLCIEN 183

QY 223 SVMTSFSK 230
DB 184 SFMTSFSK 191

RESULT 5
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATUE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-998-831-13
Query Match 68.0%; Score 843; DB 4; Length 191;
Best Local Similarity 83.0%; Pred. No. 5.6e-87;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;
QY 43 HTHHHTHQPQVPLVHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFRFLSSRLQD 102
DB 4 HHHHTHQPQVPLVHLVALNTPLSGMRGIRGADFCQQAARAVGLSGTFRFLSSRLQD 63
QY 103 LYSIVRRADRTGVPVNLRLDEVLPFSWEALFSGSEGOLKPGARIFSGDRVLOHPAWPR 162
DB 64 LYSIVRRADRGSPVIVNLKDEVLPFSWEALFSGSEGOLKPGARIFSGDRVLOHPAWPQ 123
QY 163 KSVYHSDPSGRLTDSYCYETWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLICN 222
DB 124 KSVYHSDPSGRLMESYCYETWTEATGTAIGQASSLLGRLLLEQEAASCHNSYIVLICIEN 183
QY 223 SVMTSFSK 230
DB 184 SFMTSFSK 191
RESULT 6
US-08-159-784-2
Sequence 2, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2
Query Match 67.7%; Score 839; DB 1; Length 195;
Best Local Similarity 80.2%; Pred. No. 1.6e-86;
Matches 158; Conservative 19; Mismatches 16; Indels 4; Gaps 1;
QY 34 PARPTGGVHTHTHQPQVPLVHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFR 93

DB 3 PARPT----LSLANTHQDFQPVPLVHLVALNTPLSGMRGIRGADFCQQAARAVGLSGTFR 58
QY 94 AFLSSRLQDLYSIVRRADRTGVPVNLRLDEVLPFSWEALFSGSEGOLKPGARIFSGDRD 153
DB 59 AFLSSRLQDLYSIVRRADRGSPVIVNLKDEVLPFSWEALFSGSEGOLKPGARIFSGDRD 118
QY 154 VLOHPAWPRKSVYHSDPSGRLTDSYCYETWTEAPAATGQASSLLAGRLLEQEAASCRH 213
DB 119 VLRHPAWPQKSVYHSDPSGRLMESYCYETWTEATGTAIGQASSLLGRLLLEQEAASCHN 178
QY 214 AFVVLICIENSFMTSFSK 230
DB 179 SYVILICIENSFMTSFSK 195
RESULT 7
US-09-561-500-14
Sequence 14, Application US/09561500
Patent No. 6342219
GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Brekken
TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002500
CURRENT APPLICATION NUMBER: US/09/561,500
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-561-500-14
Query Match 67.4%; Score 835; DB 4; Length 182;
Best Local Similarity 85.1%; Pred. No. 4.1e-86;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 47 HTHQDFQVPLVHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 106
DB 1 HSHRDFQVPLVHLVALNSPGLSGMRGIRGADFCQQAARAVGLAGTFRFLSSRLQDLYSI 60
QY 107 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGOLKPGARIFSGDRVLOHPAWPRKSVW 166
DB 61 VRRADRAAVPIVNLKDELLFSPWEALFSGSEGOLKPGARIFSGDKVLRHPTWPQKSVW 120
QY 167 HGSDDPSGRLTDSYCYETWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 226
DB 121 HGSDDPNGRRLTESYCYETWTEAPAATGQASSLLGRLLGQSAASCHAYIVLICIENSFMT 180
QY 227 S 227
DB 181 A 181
RESULT 8
US-09-561-108-14
Sequence 14, Application US/09561108
Patent No. 6342221
GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Brekken
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/561,108
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432

```

1 TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
2
3 FILE REFERENCE: 4001-002892
4
5 CURRENT APPLICATION NUMBER: US/09/561,499
6
7 CURRENT FILING DATE: 2000-04-28
8
9 PRIOR APPLICATION NUMBER: 60/131,432
10
11 PRIOR FILING DATE: 1999-04-28
12
13 NUMBER OF SEQ ID NOS: 44
14
15 SOFTWARE: Patentin Ver. 2.0
16
17 SEQ ID NO 14
18
19 LENGTH: 182
20
21 TYPE: PRT
22
23 ORGANISM: Artificial Sequence
24
25 FEATURE:
26
27 OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
28
29 OTHER INFORMATION: PEPTIDE
30
31 US-09-561-499-14
32
33 Query Match 67.4%; Score 835; DB 4; Length 182;
34 Best Local Similarity 85.1%; Pred. No. 4.1e-86;

```

Query Match 67.4%; Score 835; DB 4; Length 182;
Best Local Similarity 85.1%; Pred. No. 4.1e-86;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY	47	HTHQPOLVHLVVALNSQPQGMGIRGADPQCQQAARAGLACTFRAFLSSRLQDLYSI	106
Db	1	HSHRDPQPVHLVHLVALNSPLSGMGRGIRGADPQCQQAARAVGLACTFRAFLSSRLQDLYSI	60
QY	107	VERADRTGVPVNLREDEVLPFSWEALFSGSGQIKPGARIFSDGDRVLOHPNAPRKSVM	166
Db	61	VERADRAAPVIVNLKADLLFPWSALFSGSGPUPKGARIFSDGKDVLRHPTWPPQKSV	120
QY	167	HGSDPSGRLTDSYCETWRTTAPANTQAOSLLAGRLLEQFAASCRHAFVILCIENSVM	226
Db	121	HGSDPNGRRLTYESCETWRTTAPANTQAOSLLGRLLGQSRASCHHIVILCIENSFMT	180
QY	227	S	227
Db	181	A	181

```

RESULT 12
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. 6576941
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-998-831-14

```

	Query Match	67.4%	Score 835	DB 4	Length 182
	Best Local Similarity	85.1%	Pred. No. 4.4e-86		
	Matches 154	Conservative 14	Mismatches 13	Indels 0	Gaps 0
QY	47	HTHQDFQLVHLVVALNSPQCGMGIGKGAQFCQFQQAARAAGLAGTTPRAFLLSRLQDLVSI	106		
DB	1	LSHGRDFQPVHLVVALNSPLSGMGKIGKGAQFCQFQQAARAVLAGTTPRAFLLSRLQDLVSI	60		
QY	107	VRARDTGVPVVNLRDEVLPFSWEALPSGSEGOLKPGARIFSPDGRDVLQHPAWPRKSVW	166		
DB	61	VRARADRAAVPVLNKLDELLPFSWEALPSGSEGOLKPGARIFSPDGKDVLRHTPTWPKQSVW	120		
QY	167	HGSDPSGRLLTDSYCETWRTAPAAATQCASSLLAGRLLEQEAASCHAFVVLICENSVMY	226		
DB	121	HGSDPNGRLLTESYCETWRTAPASATQCASSLIGRLLGQSAASCHHAYIVLICENSFMT	180		
QY	227	S 227			
b	181	A 181			

```

RESULT 13
US-09-206-059-2
; Sequence 2, Application US/09206059
; Patent No. 6201104
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas
; APPLICANT: Sim, Kim Lee
; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
; TITLE OF INVENTION: Proteins and Methods of Use
; FILE REFERENCE: 05213-0370
;

```

```

; CURRENT APPLICATION NUMBER: US/09/206,059
; CURRENT FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-206-059-2

Query Match          67.4%; Score 835; DB 3; Length 183;
Best Local Similarity 85.1%; Pred. No. 4.2e-86;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

47 HTHODFOLVHLHVALNSPQGGMRGTRGADFOCFQOARAAAGLGFRAFLSSRLQDLYSI 106

```

```

1 HSHRDFQVFLHVLVALNSPLSGGMRGIRGADFQCFCQARAVGLAGTFRAFLSRLLQDLYSI 60
107 VRRADRTGVFWNLRDVFLPPSWEALFSSGEGKLPGARIFSFDCGRDVLQHPAPKXSVW 166
61 VRRADRAAVFVNLKDELLFPPSWEALFSSGEGPLKGARIFSFDCGKDVLRHPTWFKXSVW 120
167 HGSDPSGRRLLTDCYCTWRTTEAPAAQTQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 226
121 HGSDPENGRRLLTESYCTWRTTEAPSATQASSLLGGRLLLQSAASCHHAYIVLCIENSFWT 180
QY 227 S 227
Db 181 A 181
RESULT 14
US-09-315-689-5
; Sequence 5, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Anti-angiogenic Endostatin Compositions
; FILE REFERENCE: 052113-0229
; CURRENT APPLICATION NUMBER: US/09/315 689

```

```

; ORGANISM: HOMO SAPIENS
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-5

Query Match      65.9%; Score 817; DB 4; Length 178;
Best Local Similarity 85.9%; Pred. No. 4,3e-84;
Matches 152; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

51  DFQLVLHLVALNSPQGMGRINGADDFQCFQARAAAGLAGTTRAFLLSSRLQDLYSIVRRA 110
      1  DFQVLVLHLVALNSFLSGMGRINGADDFQCFQARAVGAGTTRAFLLSSRLQDLYSIVRRA 60

111  DRTCGPVNVLRDLVLPSPSEALFSGSEGLKFGARIFSGDRDLVQLQHPAPWPKRSVWHGSD 170
      61  DRAAVPTVNVKDBLLFPSEALFSGSEGLKFGARIFSGDKDLVLRHPTWPKSVWHGSD 120
      DB

```

[illegible]

;; TITLE OF INVENTION: CARRIER/DNA COMPLEXES CONTAINING DNA
;; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
;; TITLE OF INVENTION: THERAPY
;; NUMBER OF SEQUENCES: 43
;; CORRESPONDENCE ADDRESS:
;; ADDRESS: Connolly, Bove, Lodge, & Hutz
;; STREET: 1220 Market Street, P.O. Box 2207
;; CITY: Wilmington
;; STATE: Delaware
;; COUNTRY: U.S.A.
;; ZIP: 19899
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/985,526
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/608,845
;; FILING DATE: 16-JUL-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McMorrow Jr., Robert G
;; TELEPHONE: (302) 658-9141
;; TELEFAX: (302) 658-5613
;; INFORMATION FOR SEQ ID NO: 36:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 185 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
US-08-985-526-36

Query Match 63.8%; Score 790; DB 3; Length 185;
Best Local Similarity 81.1%; Pred. No. 5e-81;
Matches 150; Conservative 16; Mismatches 17; Indels 2; Gaps 2;
QY 47 HTHQDFQLVHLVAINSPQCGMGIQADPQCFQQAARAGLAGTFRATLSSRLQDLYSI 106
DB 2 HTHQDFQVHLVHLVAINSPQCGMGIQADPQCFQQAARAGLAGTFRATLSSRLQDLYSI 60
QY 107 VRRADRTGVPVY-NLRDEVLPFSWEALFSGSEGQKPGARIFSGDGRDVLQHPAPRKSV 165
DB 61 VRRADRGVPIVQNL-RDEVLPFSWDSLFSGSQGLQFGARIFSGDGRDVLQHPAPRKSV 120
QY 166 WHGSDPSGRRLTDSYCTETWTEAPATCAQASLLAGRLLEQEAASCHAFVYLCIENSVM 225
DB 121 WHGSDPSGRRLMESYCTETWTEATTGATCAQASLLSRLLEQEAASCHDSYVILCIENSFM 180
QY 226 TSFSK 230
DB 181 TSFSR 185

Search completed: March 26, 2004, 13:39:22
Job time : 20.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:38:08 ; Search time 41.6667 Seconds

(without alignments)
1444.373 Million cell updates/sec

Title: US-09-938-391-2

Perfect score: 1239
Sequence: 1 PWRADDILAGPRLDPPQY.....CRHAFVVLICIENSVMTSFSK 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:**

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	100.0	230	10	US-09-938-391-2
2	1019	82.2	682	15	US-10-264-049-3010
3	1019	82.2	1516	14	US-10-060-036-166
4	1019	82.2	1516	15	US-10-431-642-3
5	1003	81.0	684	10	US-09-961-403-5
6	966	78.0	184	10	US-09-938-391-4
7	959	77.4	184	14	US-10-131-241-49
8	959	77.4	184	14	US-10-292-418-35
9	843	68.0	191	9	US-09-998-831-13
10	843	68.0	191	14	US-10-373-561-13
11	835	67.4	181	14	US-10-131-241-55
12	835	67.4	182	9	US-09-998-831-14
13	835	67.4	182	14	US-10-131-241-54
14	835	67.4	182	14	US-10-042-347-3
15	835	67.4	182	14	US-10-373-561-14

Sequence 2, Appli
Sequence 1, Appli
Sequence 52, Appli
Sequence 4, Appli
Sequence 56, Appli
Sequence 47, Appli
Sequence 3, Appli
Sequence 46, Appli
Sequence 18, Appli
Sequence 71, Appli
Sequence 162, Appli
Sequence 60, Appli
Sequence 5, Appli
Sequence 57, Appli
Sequence 36, Appli
Sequence 164, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 29, Appli
Sequence 13097, A
Sequence 32, Appli
Sequence 3749, Ap
Sequence 8477, Ap
Sequence 42857, A
Sequence 41375, A
Sequence 32, Appli
Sequence 3, Appli
Sequence 4, Appli

16 835 67.4 183 9 US-09-873-676-2
17 835 67.4 183 13 US-10-080-797-1
18 835 67.4 183 14 US-10-131-241-52
19 835 67.4 183 14 US-10-292-418-4
20 834 67.3 180 14 US-10-131-241-56
21 831 67.1 180 14 US-10-131-241-47
22 831 67.1 207 13 US-10-080-797-3
23 828 66.8 184 14 US-10-131-241-46
24 828 66.8 184 14 US-10-292-418-18
25 828 66.8 207 14 US-10-422-934-71
26 822 66.3 184 12 US-10-210-172-162
27 817 65.9 178 14 US-10-131-241-60
28 817 65.9 178 14 US-10-042-347-5
29 817 65.9 179 14 US-10-131-241-57
30 790 63.8 185 13 US-10-036-869-36
31 597.5 48.2 160 12 US-10-210-172-164
32 504 40.7 160 12 US-10-210-172-166
33 302 24.4 63 9 US-09-822-540A-1
34 161 13.0 31 9 US-09-822-540A-2
35 103 8.3 24 15 US-10-016-569A-29
36 103 8.3 24 15 US-10-308-644-29
37 100.5 8.1 739 14 US-10-156-761-13097
38 97 7.8 16 9 US-09-766-412-32
39 91 7.3 431 15 US-10-108-260A-3749
40 91 7.3 6239 14 US-10-156-761-8477
41 89.5 7.2 437 12 US-10-425-114-42857
42 89 7.2 517 12 US-10-425-114-41375
43 88 7.1 3436 14 US-10-132-134-32
44 87.5 7.1 816 14 US-10-207-706-3
45 87 7.0 19725 15 US-10-084-846A-4

ALIGNMENTS

RESULT 1

US-09-938-391-2
; Sequence 2, Application US/09938391
; Publication No. US20030158099A1
; GENERAL INFORMATION:
; APPLICANT: Tong, et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING
; DISORDERS INVOLVING ANGIOGENESIS
; FILE REFERENCE: PC10790A
; CURRENT APPLICATION NUMBER: US/09/938,391
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO. 2
; LENGTH: 230
; TYPE: PRT
; ORGANISM: CANINE PRO ENDOSTATIN AMINO ACID SEQUENCE
US-09-938-391-2

Query Match 100.0%; Score 1239; DB 10; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWRADDILAGPRLDPPQYPCGPHGSGYVHFQAPRTGGPVHTHTHQQDFQLVLHLVA 60
Db 1 PWRADDILAGPRLDPPQYPCGPHGSGYVHFQAPRTGGPVHTHTHQQDFQLVLHLVA 60
QY 61 LNSPQGGNMGIRGADFCFQQAAGLAGTFFAFLSSRLQDLYSIVRRADRTGVPVNL 120
Db 61 LNSPQGGNMGIRGADFCFQQAAGLAGTFFAFLSSRLQDLYSIVRRADRTGVPVNL 120
QY 121 RDEVLPFSPWEALFSGSEGOLKFGARIFSDGRDVLQHPANPRKSVWHSQSGRLTDSY 180
Db 121 RDEVLPFSPWEALFSGSEGOLKFGARIFSDGRDVLQHPANPRKSVWHSQSGRLTDSY 180
QY 181 CETWRTAAPAAQQAASSLLAGRLLEQEAASCRHAFVLCIENSVMTSFSK 230
Db 181 CETWRTAAPAAQQAASSLLAGRLLEQEAASCRHAFVLCIENSVMTSFSK 230

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-166

Query Match      82.2%; Score 1019; DB 14; Length 1516;
Best Local Similarity 82.4%; Pred. No. 5.9e-95;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;

QY 1 PWRADDILAGPPRLDPOQYPGAPHHGVSVHFQPARPTGGPVHTHTHODFQVLVHLVA 60
DB 1292 PWRADDILASPPRLPEPQYPGAPHHSSVHLRPARPTSP-----ASHRDFQPVHLVA 1347
QY 61 LNSFPQGGMRGIRGADFCQFQOARAAAGLAGTFRAPLSRLQDLYSIVRRADRTGVPVNL 120
DB 1348 LNSPLSGMRGIRGADFCQFQOARAVAGLAGTFRAPLSRLQDLYSIVRRADRAAVPIVNL 1407
QY 121 RDEVLPFSWEALFSGSEGOLKPGARIFSDGRDVLQHPAPRKSVWHSPPSGRRRLTDSY 180
DB 1408 KDBLLFPSWEALFSGSEGOLKPGARIFSDGVKDLRHPTWPKSVWHSPPNGRRRLTESY 1467
QY 181 CETWRTAPATGAQSSLLAGLLQEAAASCRHAFVVLCIENSVMTS 227
DB 1468 CETWRTAPATGAQSSLLAGLLQEAAASCHHAYIVLCIENSPMTA 1514

RESULT 4
US-10-431-642-3
; Sequence 3, Application US/10431642
; Publication No. US20040009920A1
; GENERAL INFORMATION:
; APPLICANT: Ruoiaht, Erkki
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING
; FILE REFERENCE: BURHAM, 008CIP
; CURRENT APPLICATION NUMBER: US/10/431,642
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 10/005,171
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/331,357
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1516
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-431-642-3

Query Match      82.2%; Score 1019; DB 15; Length 1516;
Best Local Similarity 82.4%; Pred. No. 5.9e-95;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;

QY 1 PWRADDILAGPPRLDPOQYPGAPHHGVSVHFQPARPTGGPVHTHTHODFQVLVHLVA 60
DB 1292 PWRADDILASPPRLPEPQYPGAPHHSSVHLRPARPTSP-----ASHRDFQPVHLVA 1347
QY 61 LNSFPQGGMRGIRGADFCQFQOARAAAGLAGTFRAPLSRLQDLYSIVRRADRTGVPVNL 120
DB 1348 LNSPLSGMRGIRGADFCQFQOARAVAGLAGTFRAPLSRLQDLYSIVRRADRAAVPIVNL 1407
QY 121 RDEVLPFSWEALFSGSEGOLKPGARIFSDGRDVLQHPAPRKSVWHSPPSGRRRLTDSY 180
DB 1408 KDBLLFPSWEALFSGSEGOLKPGARIFSDGVKDLRHPTWPKSVWHSPPNGRRRLTESY 1467
QY 181 CETWRTAPATGAQSSLLAGLLQEAAASCRHAFVVLCIENSVMTS 227
DB 1468 CETWRTAPATGAQSSLLAGLLQEAAASCHHAYIVLCIENSPMTA 1514

RESULT 5
US-09-961-403-5
; Sequence 5, Application US/09961403
; Publication No. US20030077589A1
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-264-049-3010

Query Match      82.2%; Score 1019; DB 15; Length 682;
Best Local Similarity 82.4%; Pred. No. 2.1e-95;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;

QY 1 PWRADDILAGPPRLDPOQYPGAPHHGVSVHFQPARPTGGPVHTHTHODFQVLVHLVA 60
DB 458 PWRADDILASPPRLPEPQYPGAPHHSSVHLRPARPTSP-----ASHRDFQPVHLVA 513
QY 61 LNSFPQGGMRGIRGADFCQFQOARAAAGLAGTFRAPLSRLQDLYSIVRRADRTGVPVNL 120
DB 514 LNSPLSGMRGIRGADFCQFQOARAVAGLAGTFRAPLSRLQDLYSIVRRADRAAVPIVNL 573
QY 121 RDEVLPFSWEALFSGSEGOLKPGARIFSDGRDVLQHPAPRKSVWHSPPSGRRRLTDSY 180
DB 574 KDBLLFPSWEALFSGSEGOLKPGARIFSDGVKDLRHPTWPKSVWHSPPNGRRRLTESY 633
QY 181 CETWRTAPATGAQSSLLAGLLQEAAASCRHAFVVLCIENSVMTS 227
DB 634 CETWRTAPATGAQSSLLAGLLQEAAASCHHAYIVLCIENSPMTA 680

RESULT 3
US-10-060-036-166
; Sequence 166, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yuguu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 21021.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166
; LENGTH: 1516
```



```

; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 684
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-961-403-5

Query Match      81.0%; Score 1003; DB 10; Length 684;
Best Local Similarity 81.5%; Pred. No. 9.1e-94;
Matches 185; Conservative 15; Mismatches 23; Indels 4; Gaps 1;

QY 1 PWRADDILAGPRLLDQPIPGAPHHGSHYHFOPARPTGPFVTHTHODPOLVHLVA 60
DB 460 PWRADDILASPPGLPEPQPVPGPHSSYVHCGPARPTGPP-----AHSHRDPQPVHLVA 515
QY 61 LNSPQGGMRGIRGADFCQQAARAGLAGTFFAFLSSRLQDLYSIVREADRTGVPVNL 120
DB 516 LNSPLSGGMRGIRGADFCQQAARAGLAGTFFAFLSSRLQDLYSIVREADRAAVPIVNL 575
QY 121 RDEVLPFSWEALFSGSEGQKPGARIFSDGRVLOHPAMPKSVWHGSDPSGRLLTDSY 180
DB 576 KDELFPFSWEALFSGSEGQKPGARIFSDGKDLVLSHPTWPKSVWHGSDPNRRLTDSY 635
QY 181 CETWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
DB 636 CETWTEAPSATQASSLLGRLGQSAASCHAHYIVLCIENSFWTA 682

RESULT 6
US-09-938-391-4
; Sequence 4, Application US/09938391
; Publication No. US2003015809A1
; GENERAL INFORMATION:
; APPLICANT: Tong, et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING
; DISORDERS INVOLVING ANGIOGENESIS
; FILE REFERENCE: PCL0790A
; CURRENT APPLICATION NUMBER: US/09/938,391
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 184
; TYPE: PRT
; ORGANISM: CANINE ENDOSTATIN AMINO ACID SEQUENCE
; US-09-938-391-4

Query Match      78.0%; Score 966; DB 10; Length 184;
Best Local Similarity 100.0%; Pred. No. 1e-90;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 HTHQDFQLVHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFFAFLSSRLQDLYSI 106
DB 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFFAFLSSRLQDLYSI 60
QY 107 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGQKPGARIFSDGRVLOHPAMPKSVW 166
DB 61 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGQKPGARIFSDGRVLOHPAMPKSVW 120
QY 167 HGSDFSGRRLTDSYCEWTWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 226
DB 121 HGSDFSGRRLTDSYCEWTWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180

RESULT 7
US-10-131-241-49
; Sequence 49, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 49
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Canine sp.
; US-10-131-241-49

Query Match      77.4%; Score 959; DB 14; Length 184;
Best Local Similarity 99.5%; Pred. No. 5.3e-90;
Matches 183; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 47 HTHQDFQLVHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFFAFLSSRLQDLYSI 106
DB 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFFAFLSSRLQDLYSI 60
QY 107 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGQKPGARIFSDGRVLOHPAMPKSVW 166
DB 61 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGQKPGARIFSDGRVLOHPAMPKSVW 120
QY 167 HGSDFSGRRLTDSYCEWTWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 226
DB 121 HGSDFSGRRLTDSYCEWTWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180

RESULT 8
US-10-292-418-35
; Sequence 35, Application US/10292418
; Publication No. US20030139365A1
; GENERAL INFORMATION:
; APPLICANT: Li, Yue
; APPLICANT: Gillies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006C1
; CURRENT APPLICATION NUMBER: US/10/292,418
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/383,315
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 60/097,883
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54
```

SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Canis familiaris
US-10-292-418-35

Query Match 77.4%; Score 959; DB 14; Length 184;
Best Local Similarity 99.8%; Pred. No. 5.3e-90;
Matches 183; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 47 HTHTDQFVLHVALNSPQGMGIRGADFCFQQAAGLAGTFFRAFLSSRLQDIYSI 106
DB 1 HTHTDQFVLHVALNSPQGMGIRGADFCFQQAAGLAGTFFRAFLSSRLQDIYSI 60
QY 107 VRRADRTGVPVNLRLDEVLPFSSWEALFSGSQQLKPGARIFSPGDRDVLQHPAMPKRSVW 166
DB 61 VRRADRTGVPVNLRLDEVLPFSSWEALFSGSQQLKPGARIFSPGDRDVLQHPAMPKRSVW 120
QY 167 HGSDDPSGRRLTDSYCYETWRTTEAPATGQASSLLAGRLLEQBAASCRHAFVVLVCIENTSVMT 226
DB 121 HGSDDPSGRRLTDSYCYETWRTTEAPATGQASSLLAGRLLEQBAASCRHAFVVLVCIENTSVMT 180
QY 227 SFTSFSK 230
DB 181 SFTSFSK 191

RESULT 9
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. US20020119153A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brecken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; FILE REFERENCE: 4001.002584
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/998,831
; PRIOR FILING DATE: 2001-11-30
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-998-831-13

Query Match 68.0%; Score 843; DB 9; Length 191;
Best Local Similarity 83.0%; Pred. No. 4.3e-78;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;
QY 43 HTHTHQQDFQLVHLVALNSPQGMGIRGADFCFQQAAGLAGTFFRAFLSSRLQD 102
DB 4 HHHHTHQQDFQLVHLVALNSPQGMGIRGADFCFQQAAGLAGTFFRAFLSSRLQD 63
QY 103 LYSIVRADRTGVPVNLRLDEVLPFSSWEALFSGSQQLKPGARIFSPGDRDVLQHPAMPK 162
DB 64 LYSIVRADRTGVPVNLRLDEVLPFSSWEALFSGSQQLKPGARIFSPGDRDVLQHPAMPK 123
QY 163 KSVWHGSDPSGRRLTDSYCYETWRTTEAPATGQASSLLAGRLLEQBAASCRHAFVVLVCIENT 222
DB 124 KSVWHGSDPSGRRLTDSYCYETWRTTEAPATGQASSLLAGRLLEQBAASCRHAFVVLVCIENT 183
QY 223 SVMTSFSK 230
DB 184 SFTSFSK 191

RESULT 10
US-10-373-561-13
; Sequence 13, Application US/10373561
; Publication No. US20030175276A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brecken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-10-373-561-13

Query Match 88.0%; Score 843; DB 14; Length 191;
Best Local Similarity 83.0%; Pred. No. 4.3e-78;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;
QY 43 HTHTHQQDFQLVHLVALNSPQGMGIRGADFCFQQAAGLAGTFFRAFLSSRLQD 102
DB 4 HHHHTHQQDFQLVHLVALNSPQGMGIRGADFCFQQAAGLAGTFFRAFLSSRLQD 63
QY 103 LYSIVRADRTGVPVNLRLDEVLPFSSWEALFSGSQQLKPGARIFSPGDRDVLQHPAMPK 162
DB 64 LYSIVRADRTGVPVNLRLDEVLPFSSWEALFSGSQQLKPGARIFSPGDRDVLQHPAMPK 123
QY 163 KSVWHGSDPSGRRLTDSYCYETWRTTEAPATGQASSLLAGRLLEQBAASCRHAFVVLVCIENT 222
DB 124 KSVWHGSDPSGRRLTDSYCYETWRTTEAPATGQASSLLAGRLLEQBAASCRHAFVVLVCIENT 183
QY 223 SVMTSFSK 230
DB 184 SFTSFSK 191

RESULT 11
US-10-131-241-55
; Sequence 55, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US/09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US/09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US/06/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-55

Query Match 67.4%; Score 835; DB 14; Length 181;

Best Local Similarity 85.1%; Pred. No. 2.7e-77; Indels 0; Gaps 0;
Matches 154; Conservative 14; Mismatches 13;
QY 47 HTHQDFQVLHLVALNSPPGGMGIRGADFCFQQAARAGLAGTFRFLSSRLQDLYSI 106
DB 1 HSHRDFQVLHLVALNSPLSGGMGIRGADFCFQQAARAVGLAGTFRFLSSRLQDLYSI 60
QY 107 VRRADRTGVVNNLRDEVLPFSWEALFSGSEGQKPGARIFFSGRDVLOHPAPRKSVW 166
DB 61 VRRADRAAVPIVNLKDELLFSPWEALFSGSEGQKPGARIFFSGKDVLRHPTWPKSVW 120
QY 167 HGSDDPSGRRLTDSYCTWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 226
DB 121 HGSDDPNRRLTESYCTWTEAPATGQASSLLGRLGQSAASHYIIVLCIENSFMT 180
QY 227 S 227
DB 181 A 181

RESULT 12
US-09-998-831-14
Sequence 14, Application US/09998831
Patent No. US20020119153A1
GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Brekken
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/998,831
CURRENT FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 09/561,108
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-998-831-14

Query Match 67.4%; Score 835; DB 9; Length 182;
Best Local Similarity 85.1%; Pred. No. 2.7e-77;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 47 HTHQDFQVLHLVALNSPPGGMGIRGADFCFQQAARAGLAGTFRFLSSRLQDLYSI 106
DB 1 HSHRDFQVLHLVALNSPLSGGMGIRGADFCFQQAARAVGLAGTFRFLSSRLQDLYSI 60
QY 107 VRRADRTGVVNNLRDEVLPFSWEALFSGSEGQKPGARIFFSGRDVLOHPAPRKSVW 166
DB 61 VRRADRAAVPIVNLKDELLFSPWEALFSGSEGQKPGARIFFSGKDVLRHPTWPKSVW 120
QY 167 HGSDDPSGRRLTDSYCTWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 226
DB 121 HGSDDPNRRLTESYCTWTEAPATGQASSLLGRLGQSAASHYIIVLCIENSFMT 180
QY 227 S 227
DB 181 A 181

RESULT 13
US-10-131-241-54
Sequence 54, Application US/10131241
Publication No. US20030012792A1
GENERAL INFORMATION:
APPLICANT: Holaday, John W.
APPLICANT: Fortier, Anne H.

TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
FILE REFERENCE: 05213-0344 43170-271565
CURRENT APPLICATION NUMBER: US/10/131,241
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US 09/413,049
PRIOR FILING DATE: 1999-10-06
PRIOR APPLICATION NUMBER: US 09/316,802
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: US 60/086,586
PRIOR FILING DATE: 1998-05-22
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patentin version 3.1
SEQ ID NO 54
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
US-10-131-241-54

Query Match 67.4%; Score 835; DB 14; Length 182;
Best Local Similarity 85.1%; Pred. No. 2.7e-77;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 47 HTHQDFQVLHLVALNSPPGGMGIRGADFCFQQAARAGLAGTFRFLSSRLQDLYSI 106
DB 1 HSHRDFQVLHLVALNSPLSGGMGIRGADFCFQQAARAVGLAGTFRFLSSRLQDLYSI 60
QY 107 VRRADRTGVVNNLRDEVLPFSWEALFSGSEGQKPGARIFFSGRDVLOHPAPRKSVW 166
DB 61 VRRADRAAVPIVNLKDELLFSPWEALFSGSEGQKPGARIFFSGKDVLRHPTWPKSVW 120
QY 167 HGSDDPSGRRLTDSYCTWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 226
DB 121 HGSDDPNRRLTESYCTWTEAPATGQASSLLGRLGQSAASHYIIVLCIENSFMT 180
QY 227 S 227
DB 181 A 181

RESULT 14
US-10-042-347-3
Sequence 3, Application US/10042347
Publication No. US20030114370A1
GENERAL INFORMATION:
APPLICANT: O'Reilly, Michael S.
APPLICANT: Folkman, M. Judah
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide F
TITLE OF INVENTION: Thereof
FILE REFERENCE: 05213-0880 (43170-249874)
CURRENT APPLICATION NUMBER: US/10/042,347
CURRENT FILING DATE: 2002-01-11
PRIOR APPLICATION NUMBER: US 09/315,689
PRIOR FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: US 60/106,343
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: US 09/154,302
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: US 08/740,168
PRIOR FILING DATE: 1996-10-22
PRIOR APPLICATION NUMBER: US 60/005,835
PRIOR FILING DATE: 1995-10-23
PRIOR APPLICATION NUMBER: US 60/023,070
PRIOR FILING DATE: 1996-08-02
PRIOR APPLICATION NUMBER: US 60/026,263
PRIOR FILING DATE: 1996-09-17
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin version 3.1
SEQ ID NO 3
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
US-10-042-347-3

	Query Match	67.4%;	Score 835,	DB 14;	Length 182;
	Best Local Similarity	85.1%;	Prod. No. 2,7e-77;		
	Matches 154;	Conservative	14;	Mismatches 13;	Indels 0; Gaps 0;
QY	47	HTHDDPOLVLIHLVALNSPQGGWGRGTGADFCFQQAQAAGLAGTTFRAFUSLRLODIYSI	106		
DB	1	HSRDFQPVLIHLVALNSPLSGWGRGTGADFCFQQAQAVGLAGTTFRAFUSLRLODIYSI	60		
QY	107	VRRADRTGVPMVRLREVLPPSWEALFSGSEGLKPGARIFFSGRDVLQHPAWPRKSVW	166		
DB	61	VRRADRAAVPTVNLKDELLFSSWEALFSGSEGLKPGARIFFSGKDVLRRHETWPQKSVW	120		
QY	167	HGSDPSGRRUTDSCYETWRTEAPAAATQGASSLAGELLEQPAASCRHAFVVLCTIENSVMT	236		
DB	121	HGSDPNGRRTUESCYETWRTEAPSATQGASSLGGRLLGQAASCHHAYIVLCTIENSFMT	180		
QY	227	S 227			
DB	181	A 181			

RESULT 15

```

US-10-373-561-14
; Sequence 14, Application US/10373561
; Publication No. US20030175276A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/10/373,561
; CURRENT FILING DATE: 2003-02-24
; PRIORITY APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-10-373-561-14

```

Query Match	67.4%	Score	835,	DB 14;	Length	182;		
Best Local Similarity	85.1%;	Pred.	No. 2.7e-77;					
Matches	154;	Conservative	14;	Mismatches	13;	Indels	0; Gaps	0;
QY	47	HTHODPOLVHLVALNSPPGGMGRGIRGADFCFCQQARAAAGLAGTFFRAFLSSRLQDIYSI	106					
Ddb	1	HSHRDFQFVLHLVALNSPLSGGMGRGIRGADFCFCQQARAAGLAGTFFRAFLSSRLQDIYSI	60					
QY	107	VRRADRGTGVPNALRDEVLPFPSSWEALFSGSEQLKPGARIFFSDGRDVLQHPAMPKRSVM	166					
Ddb	61	VRRADRAAVFIVNLKDPELLPFSWEALFSGSEGELKPGARIFFSPGKVLRHPTWPKSVW	120					
QY	167	HGSPDFSGRLLTDSYCETWTWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM	226					
Ddb	121	HGSPDNNGRRLLTESYCETWTWTEAPATGQASSLLGGRLLGQAASCHAYIVLCIENSEMT	180					
QY	227	S	227					
Ddb	181	A	181					

Search completed: March 26, 2004, 13:47:33
Job time : 42.6667 secs

hypothetical prote
oligopeptide bindi
hypothetical prote
probable arabinosy
conserved hypothet
VgrG protein [limpo
hypothetical prote
chitinase (EC 3.2.
hypothetical prote
hypothetical prote
novel cellular pro
beta transducin-li
hypothetical prote
DNA photolyase (im
blue-light photore
amidophosphoribosy

30 80.5 6.5 439 2 PQ0054
31 80 6.5 588 2 AC2276
32 80 6.5 624 2 B83386
33 80 6.5 1082 2 T45096
34 79.5 6.4 346 2 G83127
35 79.5 6.4 713 2 D90658
36 79.5 6.4 713 2 D85509
37 79 6.4 558 2 T30418
38 79 6.4 672 2 T36083
39 79 6.4 1283 2 T49804
40 78.5 6.3 999 2 I38547
41 78.5 6.3 1049 2 T42045
42 78 6.3 345 2 C87288
43 78 6.3 479 2 AG2726
44 78 6.3 479 2 C97508
45 78 6.3 511 2 S52622

ALIGNMENTS

RESULT 1
A53019
collagen alpha 1(XVII) chain - human (fragment)
N:Contains: endostatin
C:Species: Homo sapiens (man)
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 15-Sep-2003
C:Accession: A53019
R:Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen,
Genomics 19, 494-499, 1994
A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localiz
A:Reference number: A53019; MUID:94245237; PMID:8186291
A:Accession: A53019
A:Molecule type: mRNA
A:Residues: 1-684 <OHA>
A:Cross-references: GB:122548; NID:9348908; PIDN:AA51864.1; PID:9562794
A:Note: The cited accession number, L25548, is not in Genbank release 103
A:Note: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifted
C:Comment: prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C:Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivas
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: GDB:COL18A1
A:Cross-references: GDB:138752; OMIM:120328
A:Map position: 21q22.3-21q22.3
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly
F:1-684/Product: collagen alpha 1(XVII) chain (fragment) #status predicted <MAT>
F:1-59/Domain: collagenous (fragment) #status predicted <CO4>
F:74-115/Domain: collagenous #status predicted <CO5>
F:129-201/Domain: collagenous #status predicted <CO6>
F:212-244/Domain: collagenous #status predicted <CO7>
F:257-278/Domain: collagenous #status predicted <CO8>
F:262-264/Region: cell attachment (R-G-D) motif
F:286-340/Domain: collagenous #status predicted <CO9>
F:354-371/Domain: collagenous #status predicted <CO10>
F:502-684/Product: endostatin #status predicted <EST>
F:509-684/Region: multiplexin collagen carboxyl-terminal homologous

Query Match 81.0%; Score 1003; DB 2; Length 684;
Best Local Similarity 81.5%; Pred. No. 9.8e-84;
Matches 185; Conservative 15; Mismatches 23; Indels 4; Gaps 1;
QY 1 PWRADDILAGPRLDPPQYPGPHGSHYHFOAPRTGPGVHTHTHTQDQLVHLVA 60
DB 460 PWRADDILAGPRLDPPQYPGPHGSHYHFOAPRTGPGVHTHTHTQDQLVHLVA 515
QY 61 LNSPQGMGIRGADQCCQQAARAGLACTPRAFLSSRLQDLYSIVRRADRTGVPVNL 120
DB 516 LNSPQGMGIRGADQCCQQAARAGLACTPRAFLSSRLQDLYSIVRRADRAAIVNL 575
QY 121 RDEVLFPSEWALFSGRGLKPGARIFDFGRDVLQHPAMPFRKSVWHGSDPSGRRLTDSY 180

OM protein - protein search, using sw model
Run on: March 26, 2004, 13:33:07 ; Search time 15.5556 Seconds
(without alignments)
1422.260 Million cell updates/sec

Title: US-09-938-391-2
Perfect score: 1239
Sequence: 1 PWRADDILAGPRLDPPQY.....CRHAFVLCIENSVMTSFSK 230
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1003	81.0	684	A53019	collagen alpha 1(X
2	983.5	79.4	1774	B56101	collagen alpha 1(X
3	980.5	79.1	1315	A56101	collagen alpha 1(X
4	528.5	42.7	1388	A53317	hypothetical prote
5	369.5	29.8	650	T22002	sulfolipid biosynt
6	99	8.0	244	C45729	58K membrane-assoc
7	98	7.9	427	A53738	probable thiamin b
8	93.5	7.5	547	T44743	hypothetical prote
9	90	7.3	404	G90781	hypothetical prote
10	90	7.3	404	D85642	hypothetical prote
11	89	7.2	335	T34086	ataxin-1 - human
12	87.5	7.1	816	S46268	alginatase biosynthe
13	87	7.0	494	A83204	hypothetical prote
14	85.5	6.9	508	T36945	hypothetical prote
15	85	6.9	300	C75235	NAD(P)-arginine AD
16	85	6.9	312	A55461	hypothetical prote
17	84	6.8	426	T17336	protein-tyrosine k
18	83	6.7	1187	TVHUT2	HC-toxin synthetas
19	82.5	6.7	5232	A45086	hypothetical prote
20	82	6.6	429	T20387	DNA-binding protei
21	82	6.6	486	A41537	RhOGAP protein hom
22	82	6.6	995	H59432	polyketide synthas
23	82	6.6	6420	T30283	probable transcrip
24	81.5	6.6	190	T36950	NAD(P)-arginine AD
25	81.5	6.6	312	B55461	probable transfera
26	81.5	6.6	566	T34842	probable Vgr prote
27	81.5	6.6	633	G90704	probable Vgr prote
28	81.5	6.6	633	B85555	probable phospholi
29	81	6.5	833	T01547	

DBB DB QY DBB

576 KDLEFLPSWEALFGSGSEGLPFGARIFSFQDKVLRHETWPKSVHGSDPNGRRLTESY 635
181 CETWRTEAPATGQASSLLAGRLLEOEAAASCRHAFVLCIENSVMTS 227
636 CETWRTEAPSATGQASSLGGRLLGQAASCHAYIVLCIENSMFTA 682

RESULT 2
CS6101 collagen alpha 1(XVIII) chain precursor, long splice form - mouse
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue-specific homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: B56101
A;Molecule type: mRNA
A;Residues: 1-562 <REHI>
A;Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430
A;Experimental source: splice form clone PE17.24
A;Author: R.R.Oh, S.F.; Kanagata, Y.; Muragaki, Y.; Tammons, S.; Ooshima, A.; Olsen, B.R.
A;Submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: B56101
A;Molecule type: mRNA
A;Residues: 1-239,487-562 <REH2>
A;Cross-references: GB:U11637; NID:g618429
A;Experimental source: splice form clones PE8.1, PE19, PE15.2
A;Author: R.R.Oh, S.F.; Kanagata, Y.; Muragaki, Y.; Tammons, S.; Ooshima, A.; Olsen, B.R.

A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa repeats from Dictyostelium discoideum.
A;Reference number: A59370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 487-1512, 'F', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
A;Cross-references: EMBL:L22545
A;Author: Ri.Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A;Title: Identification of a novel collagen chain represented by extensive interruptions
A;Reference number: PN0675; MUID:94059075; PMID:8240330
A;Accession: PN0675
A;Molecule type: mRNA
A;Residues: 635-1774 <ABE>
A;Author: Ri.Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A;Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str collagen chain.
A;Reference number: A54072; MUID:94245707; PMID:8188673
A;Accession: A54072
A;Molecule type: DNA; mRNA
A;Residues: 1293-1403, 'R', 1405-1774 <REH3>
A;Cross-references: GB:I003714; NID:9487733; PIDN:AAA20657.1; PID:9487734
A;Author: Ri,O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bi.Cell 88, 277-285, 1997
A;Title: Endostatin, an endogenous inhibitor of angiogenesis and tumor growth.
A;Reference number: A58816; MUID:97160848; PMID:9008168
A;Accession: A58816
A;Molecule type: protein
A;Residues: 1591-1610 <ORE>
A;Experimental source: hemangiopoietic endothelial cell proliferation
A>Note: inhibits endothelial cell proliferation
C.Comment: Prolines and lysines at the third position of the tripeptide repeating unit (XIII) may be involved in per-lated and subsequently O-glycosylated forms of collagen alpha 1(XVIII).

C>Date: 07-Jul-1995 #sequence revision 07-Jul-1995 #text_change 15-Sep-2003
C/Accession: A53317; A53146; S28778
R/Kivirikko, S.; Heinonen, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T.
J. Biol. Chem. 269, 4773-4779, 1994
A/Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intr
A/Reference number: A53317; MUID:94148920; PMID:8105446
A/Accession: A53317
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-138 <KIV>
A/Cross-references: GB:125280
A/Note: nucleotide sequence and conceptual translation not complete
R/Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A/Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple
A/Reference number: A53146; MUID:94140817; PMID:8307960
A/Accession: A53146
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-9, 'S', '11-48, 'V', '50-94, 'A', '96-149, 'A', '151-203, 'V', '205-408, 'A', '410-569 <MUR
A/Cross-references: GB:D21230; NID:9415605; PIDN:BA04762.1; FID:01005294; FID:9460703
R/Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A/Title: Identification of a previously unknown human collagen chain, alpha1(XV), chara
A/Reference number: S28778; MUID:93066196; PMID:1279671
A/Accession: S28778
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 544-640, 'P', '642-811, 'P', '813-1252 <MYE>
C/Genetics:
A/Gene: GDB:COL15A1
A/Cross-references: GDB:132578; OMIM:120325
A/Map position: 9q21-9q22
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F:1216-1388/Region: multiplexin collagen carboxyl-terminal homologous

Query Match 42.7%; Score 528.5; DB 2; Length 1388;
Best Local Similarity 48.2%; Pred. No. 5.8e-40;
Matches 106; Conservative 30; Mismatches 63; Indels 21; Gaps 3;
QY 11 PPRLLDPQPPGAPFHGSGVYHFPQAPRTGCPVHTHTHQQDFQLVHLVALNSPQPGMR 70
Db 1190 PPALSNPHQLLP-----PPNFISSANVE-----KPALHLAALNMPFSGDIR 1232
QY 71 GIGADPQCFQQAARAGLAGTFRFLSSRLQDLYSIVRBRDSTGVVNLRLDVLFPESWE 130
Db 1233 ----ADPQCFQQAARAGLAGTFRFLSSRLQDLYSIVRBRDSTGVVNLRLDVLFPESWE 1288
QY 131 ALPSGSGQLKPGARIFSPGDRDLVCHPAWPKSVHSGSDPSGRLTDSYCYETWTEAPA 190
Db 1289 SIFSGHGGQNMHIPIYFPGDRDINTDPSPQKVIWHGSSPHGVRLVDVNCYCAWRTADTA 1348
QY 191 ATGOASSLLAGRLLEQEAASCRHAFVVLCLNENSVMTSFSK 230
Db 1349 VTGLASPLSTGKILDQKAYSCANRLIVLCLNENSVMTSFSK 1388

RESULT 5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T22002
R/White, S.
submitted to the EMBL Data Library, October 1996
A/Reference number: Z19500
A/Accession: T22002
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-650 <WIL>
A/Cross-references: EMBL:Z81079; PIDN:CA03084.1; GSPDB:GN00019; CESP:F39H11.4
A/Experimental source: clone F39H11

A/Cross-references: GB:I16998; NID:5404754; PIDN:AAA37434.1; PID:9553894
R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A/Reference number: S72450
A/Accession: S72450
A/Molecule type: mRNA
A/Residues: 28-687, 'L', '689-734, 'F', '736-751, 'R', '753-1315 <OHM>
A/Cross-references: EMBL:L22545; NID:9348968; PIDN:AAA19787.1; PID:9511298
R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A/Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A/Reference number: A58370; MUID:94240111; PMID:8183993
A/Accession: S65595
A/Molecule type: mRNA
A/Residues: 28-1315 <OHS>
A/Cross-references: EMBL:L22545
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C/Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per
C/Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
ay be useful in treating solid tumors.
C/Genetics:
A/Gene: MGI:Coll8a1
A/Cross-references: MGI:71175
A/Map position: 10:41.0
C/Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F:1-25/Domain: signal sequence #status predicted <SIG>
F:24-235/Region: thrombospondin amino-terminal homologous
F:26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
F:327-353/Domain: collagenous #status predicted <CO1>
F:364-437/Domain: collagenous #status predicted <CO2>
F:462-583/Domain: collagenous #status predicted <CO3>
F:607-689/Domain: collagenous #status predicted <CO4>
F:704-745/Domain: collagenous #status predicted <CO5>
F:759-831/Domain: collagenous #status predicted <CO6>
F:842-874/Domain: collagenous #status predicted <CO7>
F:887-910/Domain: collagenous #status predicted <CO8>
F:892-894/Region: cell attachment (R-G-D) motif
F:918-969/Domain: collagenous #status predicted <CO9>
F:983-1000/Domain: collagenous #status predicted <CO10>
F:1132-1315/Product: endostatin #status predicted <EST>
F:1139-1315/Region: multiplexin collagen carboxyl-terminal homologous
F:126-1315/Binding site: carboxylate (Asn) (covalent) #status predicted
F:172-228/Bisulfide bonds: #status predicted
F:240, 245, 1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:451, 454, 594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 79.1%; Score 980.5; DB 2; Length 1315;
Best Local Similarity 79.7%; Pred. No. 2.5e-81;
Matches 184; Conservative 19; Mismatches 23; Indels 5; Gaps 2;
QY 1 PWRADILLAPRLLDPQPPGAP-FHGSVYHFPQAPRTGCPVHTHTHQQDFQLVHLV 59
Db 1089 PWRADILLANPRLDPQPPGAPVPHHSSVHLPPARPT----LSLAHTHQQDFQVHLV 1144
QY 60 ALNSPQPGMGIRGADPQCFQQAARAGLAGTFRFLSSRLQDLYSIVRBRDSTGVVNL 119
Db 1145 ALNTPLSGMRGIRGADPQCFQQAARAVGLSGTFRFLSSRLQDLYSIVRBRDSTGVVNL 1204
QY 120 LRDELVPFSSWEALPSGSGQLKPGARIFSPGDRDLVCHPAWPKSVHSGSDPSGRLTDS 179
Db 1205 LRDELVPFSSWEALPSGSGQLKPGARIFSPGDRDLVCHPAWPKSVHSGSDPSGRLTDS 1264
QY 180 YCETWRTTGAATGQASSLLAGRLLEQEAASCRHAFVVLCLNENSVMTSFSK 230
Db 1265 YCETWRTTGAATGQASSLLAGRLLEQEAASCRHAFVVLCLNENSVMTSFSK 1315

RESULT 4
A53317
collagen alpha 1(XV) chain precursor - human
N/Alternate names: procollagen alpha 1(XV) chain
C/Species: Homo sapiens (man)

C:Genetics:
A:Gene: CESP.F39H11.4
A:Map position: 1
A:Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3
Query Match 29.8%; Score 369.5; DB 2; Length 650;
Best Local Similarity 41.7%; Pred. No. 8e-26;
Matches 80; Conservative 27; Mismatches 70; Indels 15; Gaps 6;
QY 33 QPAPPTGPGVHTHTHQQDQLVLHLVALNSPPGQMRGIRGADQCQQAARAGLAGTFF 92
DB 451 EPQDAG--VHKDR-----VHMIALSQPFSGNLGLRGADLCQCYREARAGYTTTF 501
QY 93 RAFLSRLQDLYSIVRRADTGVFVNLRLDEVLPFSWEALFSGSEGQKPGARIFSPDGR 152
DB 502 RAMLSNVQDLVRIVHSVD--FDITVNVAGHLFPFSWRSFVNGA--QVNHAKLFPSPDRH 558
QY 153 DVLCHPAWPKSVWHGSDPGRLTDSYCEWTEAPATGQASLLAGRLLEQEMAS-- 210
DB 559 DVLNDSRWPKRVWHGSKOGGIR-ASQYCDGWRDSSLSLASHISNTSIFQSSGSEK 617
QY 211 CRHAFVVLCHEN 222
DB 618 CENKLVVLGVEN 629
RESULT 6
C45729
sulfolipid biosynthesis protein sqdc - Rhodobacter sphaeroides
C:Species: Rhodobacter sphaeroides
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: C45729; S27664
R:Benning, C.; Somerville, C.R.
J. Bacteriol. 174, 6479-6487, 1992
A:Title: Identification of an operon involved in sulfolipid biosynthesis in Rhodobacter
A:Reference number: A45729; MUID:93015699; PMID:1400200
A:Accession: C45729
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <BEN>
A:Cross-references: GB:M89780; NID:g152038; PIDN:AAA73225.1; PID:g152041
C:Genetics:
A:Gene: sqdc
A:Start codon: GTG
Query Match 8.0%; Score 99; DB 2; Length 244;
Best Local Similarity 23.9%; Pred. No. 0.13;
Matches 56; Conservative 28; Mismatches 96; Indels 54; Gaps 8;
QY 14 LLDPQPYGAPHGHSYVHFOPAPPTGPGVHTHTHQQDQLVLHLVALNSPPGQMRGIR 73
DB 2 ILASPLFGADEHRPYDLLGLPLPLA-----DVDALHCAFOH--VPGRYRGE 48
QY 74 GADPQCQQAARAGLAGTFFAFLSRLQDLYSIVRRADTGVFVNLRLDEVLPFSWEALF 133
DB 49 QDPGEF---RRANLEGLSLFPEAMGRRLFLSSRAVFDGPGTLLTEAMPPECESLY 105
QY 134 SGSEGQKPGARIFSPDGRVLOHPAMPKSVWHGSDPSGRLTDSYCEWTE-- 185
DB 106 ---GVKAEAVLNF-----AEGGASLRTATGYGPGDHK-----WRVLFEDFA 147
QY 186 -----TEAPATGQASLLAGRLLEQEMASCRHAFVVLCHENSIVNTSFSK 230
DB 148 GRPIEPRVATEVHGADLAAALL---LLEKPDAGAFVHSDLLDRHDLAEVAR 198
RESULT 7
A53798
58k membrane-associated protein - rat
N:Alternate names: 58k microfilament-associated protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 24-Nov-1999
C:Accession: A53798
R:Jiang, S.H.; Huang, J.; Li, Y.; Salas, P.J.I.; Presien, N.; Carraway, C.A.C.; Carraway
J. Biol. Chem. 269, 15067-15075, 1994
A:Title: Molecular cloning and sequencing of a 58-kDa membrane- and microfilament-associ-
A:Reference number: A53798; MUID:94253065; PMID:8195143
A:Accession: A53798
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-427 <JUA>
A:Cross-references: GB:U15425; GB:S70214; NID:g562083; PIDN:AAB54085.1; PID:g562084
C:Superfamily: mammalian retrovirus gag polyprotein 1
C:Keywords: actin binding; monomer; phosphoprotein
Query Match 7.9%; Score 98; DB 2; Length 427;
Best Local Similarity 24.9%; Pred. No. 0.32;
Matches 59; Conservative 29; Mismatches 97; Indels 52; Gaps 11;
QY 5 DDILAGPRLLDPPYCAPHGHSYVHFOPAPPTGPGVHTHTHQQDQLVLHLVALNSP 64
DB 156 DLLLEBP-----PVPVTPAPREEVE-PPAP-----RLRAAPSP 191
QY 65 QPGQVRGIR-----GADPQCQQAARAGLAGTFFAFLSRLQDLYSIVRR---ADRTGVPV 117
DB 192 VAGRLGRREVAPDSTSQAPFLRQAGGQWQWPFSA--DIYNWKHNPPFPKDPVAL 248
QY 118 VNLRLDEVLF---PSW-----BALFSGSEGO---LKPGRIFSPDGRDLQ---HPAW 160
DB 249 TNLIESVLLTHQPTWDDIQQLLQALITSEKQVLEAKHVLGNGRPTLLPEIDDAF 308
QY 161 P-RKSVWHGSDPGSRRRLTDSYCEWTEAPATGQASLLAGRLLEQEMASCRHAFV 216
DB 309 PLTRPDWDTTAEGRRLRLYQLLGLAGRAARRPTNLIAQVKVQVQEAATPFAFL 365
RESULT 8
T44743
probable thiamin biosynthesis protein thicC [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C:Accession: T44743
R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A:Reference number: 222831
A:Accession: T44743
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-547 <JAM>
A:Cross-references: EMBL:AL035159; PIDN:CAA22712.1
A:Experimental source: cosmid B1450
C:Genetics:
A:Gene: thicC
C:Superfamily: thiamin biosynthesis protein thicC
Query Match 7.5%; Score 93.5; DB 2; Length 547;
Best Local Similarity 27.4%; Pred. No. 1.1;
Matches 43; Conservative 14; Mismatches 49; Indels 51; Gaps 8;
QY 19 PYCAPHGHSYVHFOPAPPTGPG-----VHTHTHQQD-----QLVLHLVA 60
DB 17 PIPGS-----SKATREVANPDGGSLSVPPFRVLSLSTGAHFDLYTSGPYTDPDAVINLTA 72
QY 61 LNSPQPGMRGIRGADPQCQQAARAGLAGTFFAFLSRLQDLYSIVRRADTGVFVNL 120
DB 73 GLPPEPGVIRD-RGTQ---LQARAGEITAEI-AFT-----ADREGMPAEV 114
QY 121 RDEVLF-----PSWEALFSGSEGQKPGARI 146
DB 115 RVEVALGRAVTPANHHHPETIEPMILGKFAVKVANI 151
RESULT 9
G90781
hypothetical protein ECs1223 [imported] - Escherichia coli (strain O157:H7, substrain I
C:Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: G90781
E;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G90781
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA034646.1; PID:g13360683; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 0509952
C;Genetics:
A;Gene: ECs1223

Query Match 7.3%; Score 90; DB 2; Length 404;
Best Local Similarity 25.5%; Pred. No. 1.6; Indels 107; Gaps 18;
Matches 73; Conservative 28; Mismatches 107; Indels 78; Gaps 18;

Qy 9 AGPP--RLDDPQYPGAPHHGVSYPHQPAPRTGGPVHTH-----THQDFQLVL-----HL 58
Db 53 AGAPVVRITDLNKAGDEVFTFSIMHKLKRPWTGDERVEGRGEDLSHADFSKINQGRHL 112
Qy 59 VALNSPQGGMRGIRGADFCFQQAAGLAGTFRAFLLSRLOLYSIVRR-----A 110
Db 113 V-----DAGGRMSQRTKFNLAASART--LLGTY-----FNDLQDQCAIVHLAARGDFVA 161
Qy 111 DRTGVPPVN-----LRDEVLPFSWEALFSGSEG-----QLKPGARIFSF-----DGR----- 152
Db 162 DDTILPTAEHPFKKIMINDVLPPTHDRHFFGGDATSPQIE-AAIDIFSIGLVNLSLFI 220
Qy 153 DVLQHPAWPRK---SVWHGSDP-----SGRLTDSYCYT-----WRTEAPATGQASSLL 199
Db 221 DEMAHPLOPVLRLSGDELHGDEPYVLYVTPRQNDWYTSKGDKNMQMVRAVRAKGF-F 279
Qy 200 AGRLEQEAASCRHAFV-----VLCIENSVMTSFSK 230
Db 280 NHPLFKGECAMRNILVRKYAGMPIRFYQGSKVLVSENN-LTATTK 324

RESULT 10
D85642
hypothetical protein Z1479 (imported) - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85642
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11204551
A;Accession: D85642
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <STO>
A;Cross-references: GB:AE005174; NID:g12514332; PIDN:AA055600.1; GSPDB:GN00145; UWGP:Z14
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1479

Query Match 7.3%; Score 90; DB 2; Length 404;
Best Local Similarity 25.5%; Pred. No. 1.6;
Matches 73; Conservative 28; Mismatches 107; Indels 78; Gaps 18;

Qy 9 AGPP--RLDDPQYPGAPHHGVSYPHQPAPRTGGPVHTH-----THQDFQLVL-----HL 58
Db 53 AGAPVVRITDLNKAGDEVFTFSIMHKLKRPWTGDERVEGRGEDLSHADFSKINQGRHL 112
Qy 59 VALNSPQGGMRGIRGADFCFQQAAGLAGTFRAFLLSRLOLYSIVRR-----A 110
Db 113 V-----DAGGRMSQRTKFNLAASART--LLGTY-----FNDLQDQCAIVHLAARGDFVA 161

Qy 111 DRTGVPPVN-----LRDEVLPFSWEALFSGSEG-----QLKPGARIFSF-----DGR----- 152
Db 162 DDTILPTAEHPFKKIMINDVLPPTHDRHFFGGDATSPQIE-AAIDIFSIGLVNLSLFI 220
Qy 153 DVLQHPAWPRK---SVWHGSDP-----SGRLTDSYCYT-----WRTEAPATGQASSLL 199
Db 221 DEMAHPLOPVLRLSGDELHGDEPYVLYVTPRQNDWYTSKGDKNMQMVRAVRAKGF-F 279
Qy 200 AGRLEQEAASCRHAFV-----VLCIENSVMTSFSK 230
Db 280 NHPLFKGECAMRNILVRKYAGMPIRFYQGSKVLVSENN-LTATTK 324

RESULT 11
T34086
hypothetical protein C06E7.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34086
R;Le, T.
A;Description: The sequence of C. elegans cosmid C06E7.
A;Reference number: Z21474
A;Accession: T34086
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-335 <LET>
A;Cross-references: EMBL:U41009; PIDN:AAA82281.1; CRSP:C06E7.4
C;Genetics:
A;Gene: CRSP:C06E7.4
A;Introns: 41/2; 83/1; 116/3; 219/1; 273/3

Query Match 7.2%; Score 89; DB 2; Length 335;
Best Local Similarity 30.3%; Pred. No. 1.6; Indels 20; Gaps 3;
Matches 20; Conservative 4; Mismatches 22; Indels 20; Gaps 3;
Qy 1 PMRADDILAGPRLDDPQYPGAPHHGVSYPHQPAPRTGGPVHTHTHQDFQLVHLVA 60
Db 71 PMNTSSISSGKRDIISPPMYSSTFHH-----HNR-HNQQLQ-----HHQV 110
Qy 61 LNSFPQ 66
Db 111 IKSFPQ 116

RESULT 12
S46268
ataxin-1 - human
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S46268
R;Banfi, S.; Servadio, A.; Chung, M.; Kwiatkowski Jr., T.J.; McCall, A.E.; Duwick, L.A
Nature Genet. 7, 513-519, 1994
A;Title: Identification and characterization of the gene causing type 1 spinocerebellar
A;Reference number: S46268; MUID:95038838; PMID:7951322
A;Accession: S46268
A;Molecule type: mRNA
A;Residues: 1-816 <BAN>
A;Cross-references: EMBL:X79204; NID:g529661; PIDN:CAA55793.1; PID:g529662
C;Comment: Spinocerebellar ataxia type 1, a dominantly inherited neurodegenerative dise
repeats in SCA1 patients.
C;Genetics:
A;Gene: GDB:SCA1
A;Cross-references: GDB:119588; OMIM:164400
A;Map position: 6p23-6p23

Query Match 7.1%; Score 87.5; DB 2; Length 816;
Best Local Similarity 22.4%; Pred. No. 6.5; Indels 65; Gaps 10;
Matches 54; Conservative 26; Mismatches 96; Indels 65; Gaps 10;
Qy 8 LAGPPRLDDPQYPGAPHHGVSYPH-----QPAPRTGGPVHTHTHQDFQLVHLV 59
Db 228 LSRAPGLITPGSPPPA-QNQYVHISSPQNTGRTASPPAIPVHLHPHQ-----TMIPHTL 282

QY. 60 ALNSPQGMGR-GIRGADFQCFQQAARAGLAGTFRATLSRRLQD-----LYSIVRAD 111
 Db 283 TLGPFQVMQVADSGSHFVPREATKAE-----SSRLQQAIOAKEVLINGEMEKSR 333
 QY 112 RTGVVPMVNLDEVLPFPMWEALFGSGSQKPGARI--FSFDGRDYLQHPAMPKRKSVWHGS 169
 Db 334 RYGAP-----SSADLGLKAGGKSVPFYSRHRVVVHPS---PSDYSSR 374
 QY 170 DPSGRRLT-----DSYCETWRTAPATGQASSLLAGRLLECEAASCCHAF 215
 Db 375 DPSGVRASVMVLPNSNTPAADLEVQAQTHREASPTSLNDKSLGLKGPCHRSYALSPTHV 434
 QY 216 V 216
 Db 435 I 435
 RESULT 13
 A83204
 alginate biosynthesis protein Alg8 PA3541 [imported] - Pseudomonas aeruginosa (strain PA
 C.Species: Pseudomonas aeruginosa
 C.Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C.Accession: A83204
 R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 A.Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 Nature 406, 959-964, 2000
 A.Reference number: AB2950; MUID:20437337; PMID:10984043
 A.Accession: A83204
 A.Status: Preliminary
 A.Molecule type: DNA
 A.Residues: 1-494 <STO>
 A.Cross-references: GB:AE004774; GB:AE0049684; PIDN:G9949684; PIDN:AAG06929.1; GSPDB:GN001
 A.Experimental source: strain PA01
 C.Genetics:
 A.Gene: PA3541

Query Match		7.0%; Score 87; DB 2; Length 494;	
Best Local Similarity	24.6%; Fred.No. 3.9;		
Matches	45; Conservative	20; Mismatches 68; Indels 50; Gaps 7;	
QY	50 QDFVLVHLVALNSPQGGMKRGADPF-----QCFOQARAGLACTFRAFLSSR	99	
DB	39 KDFILLIGAVIWRYSMGGVHURGLFVLHVVVYPYRRRVQLGSAADPSHFVLMVTSPR	98	
QY	100 LQDL-----YSIVRRADRTGVP-----VNLNRDEVLPFS-WEAL-----	132	
DB	99 IDALTAMVRSVIREAIDSGYPTTVVCISIVMSDEVIIVRSLEKMNPDPDVSLDVFRIP	158	
QY	133 FSGSEGLKPGAKRIFS-----FDGRDVLQHPEAPWPKSVWHGSDPS-GRILITDS	179	
DB	159 GTGKRDELGLAYGFRAISRHLRPDDDAVVAVIDGTVLDHGVKKVFVFKLFENVGGLTINE	218	
QY	180 YCE 182		
DB	219 PCE 221		

RESULT 14
T36945
hypothetical protein SCJ1.12 - Streptomyces coelicolor
C1:Species: Streptomyces coelicolor
C1:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C1:Accession: T36945
C1:R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21607
A:Accession: T36945
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-508 <SEE>

[illegible]

RESULT 15
 G75295
 hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: G75295
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 , M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: G75295
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-300 <WRI>
 A:Cross-references: GB:AE002058; GB:AE000513; NID:G6460059; PIDN:AAF11808.1; PID:G646001
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2255
 A:Map position: 1

Query Match	6.9%	Score 85;	DB 2;	Length 300;	
Best Local Similarity	22.1%	Pred. No. 3.2;			
Matches 54;	Conservative 17;	Mismatches 87;	Indels 86;	Gaps 10;	
23	APHGGSVV-----	HFQAPRTGPGPVHHTTHQDFQLVLHLVALNSQPQ	66		
dDb	45	APENGYVSETAGVSGLSLWLPDPHPSHAWVG-LHLHPDHADAALAEFLAAQARRA	103		
67	G-----	GMR-----	GIRGADFCQFOQARAAGL-----	88	
dDb	104	GRSHLWASVEDYLPAMPDLPALGPFVEVHRTFGGPHLRDMQVNTGALAEALTARHYLT	163		
QY	89	-AGNFRAFLSLQDLVSIYRRADRGVGVNLRDEVLPFSWEALFSGEGQKPGGARIF	147		
dDb	164	PAAPFQH--DARTEIYALTRQQVTAFTID-----	PAEAL-----	199	
QY	148	SFDGRDVLQHPAWPRKSVNWHGSDPSGRRITD--SYCETWTEAPATGTQASSLLAGRLLEQ	206		
dDb	200	--TDDELWNAAW---LAWHGETLVGLALPERSRLHAWNVLVVPVHRRQGLATALLAQ	254		
207	EAS	210			

Db 255 VARS 258

Search completed: March 26, 2004, 13:38:38
Job time : 19.5556 secs

RX MEDLINE=20400145; PubMed=10942434;
RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA Passos-Bueno M.R.;
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT and tumor growth, plays a critical role in the maintenance of retinal
RT structure and in neural tube closure.";
RL Hum. Mol. Genet. 9:2051-2058(2000).
[6]
RN VARIANT ASN-1437.
RP MEDLINE=21518361; PubMed=1160364;
RX Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RA "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT for the development of prostatic adenocarcinoma";
RL Cancer Res. 61:7375-7378(2001).
CC -!- FUNCTION: COL18A1 probably plays a major role in determining the
CC retinal structure as well as in the closure of the neural tube.
CC -!- FUNCTION: Endostatin potentially inhibits endothelial cell
CC proliferation and angiogenesis. May inhibit angiogenesis by
CC binding to the heparan sulphate proteoglycans involved in growth
CC factor signalling.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=NC-493;
CC IsoId=P39060-1; Sequence=Displayed;
CC Name=Short; Synonyms=NC1-303;
CC IsoId=P39060-2; Sequence=VSP_001155; VSP_001156;
CC -!- TISSUE SPECIFICITY: Present in multiple organs with highest levels
CC in liver, lung and kidney.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- POLYMORPHISM: There is an association between a polymorphism in
CC position 1437 and prostate cancer. Heterozygous Asn-1437
CC individuals have a 2.5 times increased chance of developing
CC prostate cancer as compared with homozygous Asp-1437 individuals.
CC -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
CC [MIM:267750]; an autosomal recessive disorder defined by the
CC occurrence of high myopia, vitreoretinal degeneration with retinal
CC detachment, macular abnormalities and occipital encephalocele.
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF018081; AAC39658.1; -
CC EMBL; AF018082; AAC39659.1; -
CC EMBL; AL163302; CAB90482.1; -
CC EMBL; L22548; AAA51864.1; -
CC EMBL; AF184060; AAF0310.1; ALT_INIT.
CC PDB; 1BNL; O2-DEC-98.
CC GlycoSuiteDB; P39060; -
CC Genew; HGNC:2195; COL18A1.
CC MIM; 120328; -
CC MIM; 267750; -
CC GO; GO:0005581; C:collagen; TAS.
CC GO; GO:0007397; P:histogenesis and organogenesis; TAS.
CC GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
CC GO; GO:0007601; P:vision; TAS.
CC GO; GO:0007601; P:vision; TAS.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR003129; TSPN.
CC Pfam; PF01391; Collagen; 7.
CC Pfam; PF02210; TSPN; 1.

DR ProDom; PD000007; Clg_helix; 1.
DR SMART; SM00382; LamG_1.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW Polymorphism; 3D-structure.
FT SIGNAL 1 23
FT CHAIN 24 1516
FT CHAIN 1334 1516
FT CHAIN 221 409
FT DOMAIN 410 516
FT DOMAIN 517 550
FT DOMAIN 551 560
FT DOMAIN 561 640
FT DOMAIN 641 664
FT DOMAIN 665 786
FT DOMAIN 787 809
FT DOMAIN 810 892
FT DOMAIN 893 906
FT DOMAIN 907 948
FT DOMAIN 949 961
FT DOMAIN 962 1034
FT DOMAIN 1035 1044
FT DOMAIN 1045 1077
FT DOMAIN 1078 1089
FT DOMAIN 1090 1111
FT DOMAIN 1112 1118
FT DOMAIN 1119 1173
FT DOMAIN 1174 1186
FT DOMAIN 1187 1204
FT DOMAIN 1205 1516
FT CARBOHYD 68 68
FT CARBOHYD 129 129
FT CARBOHYD 164 164
FT CARBOHYD 691 691
FT CARBOHYD 1329 1329
FT DISULFID 1366 1506
FT DISULFID 1468 1498
FT SITE 1095 1097
FT VARSPLIC 1 180
FT VARSPLIC 181 215
FT VARIAT 1437 1437
CONFLICT 428 428
CONFLICT 841 841
CONFLICT 877 877
CONFLICT 886 886
CONFLICT 912 912
CONFLICT 933 933
CONFLICT 975 975
CONFLICT 1064 1064
CONFLICT 1084 1084
CONFLICT 1120 1120
CONFLICT 1123 1123
CONFLICT 1126 1126
CONFLICT 1206 1206
CONFLICT 1304 1304
CONFLICT 1314 1314
CONFLICT 1324 1324
CONFLICT 1443 1443
CONFLICT 1483 1483
SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476EE76 CRC64;
Query Match 82.2%; Score 1019; DB 1; Length 1516;
Best Local Similarity 82.4%; Pred. No. 3.8e-82;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;

Qy 1 PWRADDILAGPPRLDPQYFGAPHGSHYVHPQAPRTGPGVHTHTHDPQLVHLVA 60
Db 1292 PWRADDILAGPPRLDPQYFGAPHGSHYVHLRPARPTSP-----AHSHRDFQPVHLVA 1347
Qy 61 LNSPQMGEGIGADPOCFQQAARAGLAGTFRFLSSRLQDLYSIVRADRTPGVVNL 120
Db 1348 LNSPLSGMGEGIGADPOCFQQAARAGLAGTFRFLSSRLQDLYSIVRADRAAIVVNL 1407
Qy 121 RDEVLPSWEALFSGSEGQKPCARIFSDGRDVLQHPAPWPKSVHMGSDPSGRELTDYS 180
Db 1408 KDELFPSEALFSGSEGQKPCARIFSDGRDVLQHPAPWPKSVHMGSDPSGRELTDYS 1467
Qy 181 CETWTEAPATQASLLAGRLLEQEAASCHAFVVLCIENSVMTS 227
Db 1468 CETWTEAPATQASLLAGRLLEQEAASCHAFVVLCIENSVMTS 1514

RESULT 2

CA1H_MOUSE STANDARD; PRT; 1774 AA.
ID CA1H_MOUSE STANDARD; PRT; 1774 AA.
AC P39061; Q60672; Q61437; Q62001; Q62002; Q9JK63;
DT 01-FEB-1995 (Rel. 31, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M.V., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVII collagen,
RT partial structure of the alpha 1 chain of mouse type XVII collagen,
RT alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen
RT chain.";
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=96435922; PubMed=8838808;
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT "Characterization of the mouse gene for the alpha-1 chain of type
RT XVII collagen (COL18A1) reveals that the three variant N-terminal
RT polypeptide forms are transcribed from two widely separated
RT promoters.";
RL Genomics 32:436-446(1996).
RN [3]
RP SEQUENCE OF 1-1387 FROM N.A. (ISOFORM 3).
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M.V., Pihlajaniemi T.;
RT "Alpha 1(XVII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [4]
RP SEQUENCE OF 487-1774 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94240111; PubMed=8183893;
RA Oh S.P., Kanagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT "Isolation and sequencing of cDNAs for proteins with multiple domains
RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN [5]
RP SEQUENCE OF 1591-1774 FROM N.A.
RX MEDLINE=21217748; PubMed=11321448;
RA Jia S., Zhu F., Li H., He F., Xiu R.-J.;
RT "Anticancer treatment of endostatin gene therapy by targeting tumor
RT neovasculation in C57BL mice.";
RL Clin. Hemorheol. Microcirc. 23:251-257(2000).

[6]
RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE=97160848; PubMed=9008168;
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT growth.";
RL Cell 88:277-285(1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX MEDLINE=98169382; PubMed=9501087;
RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT resolution.";
RL EMBO J. 17:1656-1664(1998).
CC -1- FUNCTION: Endostatin potentially inhibits endothelial cell
CC proliferation and angiogenesis. May inhibit angiogenesis by
CC binding to the heparan sulfate proteoglycans involved in growth
CC factor signaling.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative promoter;
CC Comment=2 isoforms, 1 (shown here) and 3, are produced by use
CC of alternative promoters;
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1; Synonyms=NC1-764;
CC IsoId=P39061-3; Sequence=Displayed;
CC Name=2; Synonyms=Long, NC1-517;
CC IsoId=P39061-1; Sequence=VSP_008303;
CC Notes=Produced by alternative splicing of isoform 1;
CC Name=3; Synonyms=Short, NC1-301;
CC IsoId=P39061-2; Sequence=VSP_001157, VSP_001158;
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; L16898; AAA37434.1; -
EMBL; U03714; AAA20657.1; -
EMBL; U03715; AAC52901.1; -
EMBL; U34606; AAC52901.1; JOINED.
EMBL; U34608; AAC52901.1; JOINED.
EMBL; U34609; AAC52901.1; JOINED.
EMBL; U34610; AAC52901.1; JOINED.
EMBL; U34611; AAC52901.1; JOINED.
EMBL; U34612; AAC52901.1; JOINED.
EMBL; U34613; AAC52901.1; JOINED.
EMBL; U03716; AAC52901.1; JOINED.
EMBL; U03718; AAC52901.1; JOINED.
EMBL; U03715; AAC52902.1; -
EMBL; U34607; AAC52902.1; JOINED.
EMBL; U34608; AAC52902.1; JOINED.
EMBL; U34609; AAC52902.1; JOINED.
EMBL; U34610; AAC52902.1; JOINED.
EMBL; U34611; AAC52902.1; JOINED.
EMBL; U34612; AAC52902.1; JOINED.
EMBL; U34613; AAC52902.1; JOINED.
EMBL; U03716; AAC52902.1; JOINED.
EMBL; U03718; AAC52902.1; JOINED.
EMBL; U03715; AAC52903.1; -
EMBL; U03716; AAC52903.1; JOINED.
EMBL; U03718; AAC52903.1; JOINED.
EMBL; U34607; AAC52903.1; JOINED.
EMBL; U34608; AAC52903.1; JOINED.

DR EMBL; U34609; AAC52903.1; JOINED.
 DR EMBL; U34610; AAC52903.1; JOINED.
 DR EMBL; U34611; AAC52903.1; JOINED.
 DR EMBL; U34612; AAC52903.1; JOINED.
 DR EMBL; U34613; AAC52903.1; JOINED.
 DR EMBL; U11636; AAC52178.1; --
 DR EMBL; U11637; AAC52179.1; --
 DR EMBL; L22545; AAA19787.1; --
 DR EMBL; AF25775; AA669009.1; --
 DR PIR; A56101; A56101.
 DR PDB; 1KOE; 16-FEB-99.
 DR PDB; 1DY0; 11-APR-00.
 DR PDB; 1DYL; 21-JAN-01.
 DR MGD; MGI:88451; Col18a1.
 DR GO; GO:0005604; C-basement membrane; IDA.
 DR GO; GO:0001525; P-angiogenesis; IMP.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008180; Collagen.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR001791; laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 8.
 DR Pfam; PF02210; TSPN; 1.
 DR ProDom; PD000007; Clg_helix; 1.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS50038; FZ; 1.
 DR EXTRACELLULAR MATRIX; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
 KW Alternative promoter usage; 3D-structure.
 FT SIGNAL 1 26
 FT CHAIN 27 1774
 FT CHAIN 1591 1774
 FT DOMAIN 345 433
 FT DOMAIN 365 482
 FT DOMAIN 27 785
 FT DOMAIN 786 812
 FT DOMAIN 813 822
 FT DOMAIN 823 896
 FT DOMAIN 897 920
 FT DOMAIN 921 1042
 FT DOMAIN 1043 1065
 FT DOMAIN 1066 1148
 FT DOMAIN 1149 1162
 FT DOMAIN 1163 1204
 FT DOMAIN 1205 1217
 FT DOMAIN 1218 1290
 FT DOMAIN 1291 1300
 FT DOMAIN 1301 1333
 FT DOMAIN 1334 1345
 FT DOMAIN 1346 1369
 FT DOMAIN 1370 1376
 FT DOMAIN 1377 1428
 FT DOMAIN 1429 1441
 FT DOMAIN 1442 1459
 FT DOMAIN 1460 1774
 FT CARBOHYD 354 354
 FT CARBOHYD 361 361
 FT CARBOHYD 585 585
 FT CARBOHYD 597 597
 FT DISULFID 1623 1763
 FT DISULFID 1725 1755
 FT SITE 1351 1353
 FT VARSPPLIC 1 459
 FT VARSPPLIC 460 486
 FT VARSPPLIC 240 486
 FT CONFLICT 1147 1147
 FT CONFLICT 1194 1194
 FT CONFLICT 1211 1211

Query Match 79.4%; Score 983.5; DB 1; Length 1774;
 Best Local Similarity 80.1%; Pred. No. 6.3e-79;
 Matches 185; Conservative 18; Mismatches 23; Indels 5; Gaps 2;
 QY 1 PWRADDILAGPPRLDPPYPGAP-HHGSVHFQFQPARPTGGPVHHTHTHDTQDFQVVLHLV 59
 DB 1548 PWRADDILANPPRLPPQYPYPGVPHHSSVHLPPARPT---LSLAHTHQDFQVVLHLV 1603
 QY 60 ALNSPQPGMRGIRGADFCFQQAARAGLAGTFRAPLSRLQDIYSIVRRADTGVVNV 119
 DB 1604 ALNTPLSGMRGIRGADFCFQQAARAGLAGTFRAPLSRLQDIYSIVRRADTGVVNV 1663
 QY 120 LRDEVLFPSEALFSGSEGQKFGARIFSDGRDVLQHPAPKSVHSGSPSGRRLTDS 179
 DB 1664 LKDEVLPSEWLSLFSGSGQQLQPGARIFSDGRDVLQHPAPKSVHSGSPSGRRLTDS 1723
 QY 180 YCETWRTEAPATGQASSILAGRLLEQEAASCHAFVVLCIENSVMTSFSK 230
 DB 1724 YCETWRTEATGQASSLLSGRLLEQEAASCHAFVVLCIENSVMTSFSK 1774
 RESULT 3
 ID CALE HUMAN STANDARD; PRT; 1388 AA.
 AC P39059.
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(XV) chain precursor.
 GN COL15A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Umbilical cord;
 RX MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Heinamaki P., Rehn M.V., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 RT exon-intron organization in the 3' region of the corresponding
 RT gene.";
 RL J. Biol. Chem. 269:4773-4779(1994).
 RN [2]
 RP SEQUENCE OF 1-569 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=94140817; PubMed=8307960;
 RA Muragaki Y., Abe N., Ninomiya Y., Olsen B.R., Ooshima A.;
 RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
 RT non-triple helical domain with a tandem repeat structure and homology
 RT to alpha 1(XVIII) collagen.";
 RL J. Biol. Chem. 269:4042-4046(1994).
 RN [3]
 RP SEQUENCE OF 544-1252 FROM N.A.
 RX MEDLINE=93066196; PubMed=1279671;
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
 RT "Identification of a previously unknown human collagen chain, alpha
 RT 1(XV), characterized by extensive interruptions in the triple-helical
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
 CC -/- TISSUE SPECIFICITY: Expressed predominantly in internal organs
 CC such as adrenal gland, pancreas and kidney.
 CC -/- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -/- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -/- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; L25286; AA58429.1; -
EMBL; D21230; BAA04762.1; -
EMBL; L01697; -; NOT_ANNOTATED_CDS.
PIR; A53317; A53317.
HSSP; P39061; 1KOB.
Genew; HGNC:2192; COL15A1.
MIM; 120325;
InterPro; IPR008160; Collagen.
InterPro; IPR008985; Cona like lec_g1.
InterPro; IPR001791; Laminin_G.
InterPro; IPR003129; TSPN.
Pfam; PF01391; Collagen; 5.
Pfam; PF02210; TSPN; 1.
SMART; SM00282; LamG; 1.
SMART; SM00210; TSPN; 1.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Cell adhesion; Collagen; Glycoprotein; Signal.
SIGNAL 1 25
CHAIN 26 1388
COLLAGEN ALPHA 1(XV) CHAIN.
TSP N-TERMINAL.
NON-HELICAL REGION 1 (NC1).
TRIPLE-HELICAL REGION 1 (COL1).
NON-HELICAL REGION 2 (NC2).
TRIPLE-HELICAL REGION 2 (COL2).
NON-HELICAL REGION 3 (NC3).
TRIPLE-HELICAL REGION 3 (COL3).
NON-HELICAL REGION 4 (NC4).
TRIPLE-HELICAL REGION 4 (COL4).
NON-HELICAL REGION 5 (NC5).
TRIPLE-HELICAL REGION 5 (COL5).
NON-HELICAL REGION 6 (NC6).
TRIPLE-HELICAL REGION 6 (COL6).
NON-HELICAL REGION 7 (NC7).
TRIPLE-HELICAL REGION 7 (COL7).
NON-HELICAL REGION 8 (NC8).
TRIPLE-HELICAL REGION 8 (COL8).
NON-HELICAL REGION 9 (NC9).
TRIPLE-HELICAL REGION 9 (COL9).
NON-HELICAL REGION 10 (NC10).
4 X TANDEM REPEATS.
1.
2.
3.
4.
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
C -> S (IN REF. 2).
D -> V (IN REF. 2).
L -> A (IN REF. 2).
P -> A (IN REF. 2).
M -> V (IN REF. 2).
R -> A (IN REF. 2).
C -> S (IN REF. 2).
10
95
95
150
204
409
1388 AA; 60822AD925A3093D CRC64;
Query Match 42.7%; Score 528.5; DB 1; Length 1388;
Best Local Similarity 48.2%; Pred. No. 8.5e-39;
Matches 106; Conservative 30; Mismatches 63; Indels 21; Gaps 3;
QY 11 PPRLLDQPYGAPHHGYSYHFQAPRTGGVTHTHQDFQVLHLVALNSPQGMVR 70
DB 1190 PPPALSNPHQLLP- - - - -PPNPISSANYE- - - - -KPAHLAALNMPFGDIR 1232
QY 71 GIRGADFQFOAARAGLAGTFRATFSLSSRLQDLYSVRRADRTGVVNLRLDVLFPSEW 130
|||||

1233 ----ADPQCFKQARAGLLSYRAFLSSHLQDLSTIVKAEYSLSPIVNLKQVLFNNWD 1288
QY 131 ALFSGSEGLKPGARIFDFDQDVLQHPAMPKRWGSDPSGRRLTDSYCYETWRTEAPA 190
1289 SIFSGGGGNNHHIPIYDFDGRDINTDPSNPQKVIWHSPPHGVRLVDNYCEAWRTADTA 1348
QY 191 ATQOASSILAGLLRLEQEAASCRHAFVVLVCIENSVMYFSK 230
1349 VTGLASPLSTGKILQKAYSCANRLVLCIENSFWTDARK 1388
DB
RESULT 4
THIC MYCLE
ID THIC MYCLE STANDARD; PRT; 547 AA.
AC Q9ZBL0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB Thiamine biosynthesis protein thic.
GN THIC OR ML0294 OR MLCB1450.28C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
SEQUENCE FROM N.A.
STRAIN=TN;
MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomsen N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: Required for the synthesis of the hydromethylpyrimidine
(HMP) moiety of thiamine (4-amino-2-methyl-5-
hydroxymethylpyrimidine) (By similarity).
CC -!- PATHWAY: Thiamine biosynthesis.
CC -!- SIMILARITY: Belongs to the thic family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; AL095159; CAA22712.1; -
EMBL; AL583918; CAC29802.1; -
PIR; T44743; T44743.
DR Leptoma; ML0294; -
DR HAMAP; MF_00089; -; 1.
DR InterPro; IPR002817; Thic.
DR Pfam; PF01964; Thic; 1.
DR ProDom; PD007049; Thic; 1.
DR TIGRFAMs; TIGR00190; thic; 1.
DR Thiamine biosynthesis; Complete proteome.
KW
SQ SEQUENCE 547 AA; 59840 MW; 797DFB09DE462D58 CRC64;
Query Match 7.5%; Score 93.5; DB 1; Length 547;
Best Local Similarity 27.4%; Pred. No. 0.87; Indels 51; Gaps 8;
Matches 43; Conservative 14; Mismatches 14;
QY 19 FYPGAPHHGYSYHFQAPRTGGP- - - - -VHTHTTHQDF- - - - -OLVHLVA 60
DB 17 PIPGS- - - - -SKAYEVANPDGPGSLRVPFRVHLSTGAFDLYDTSGPYTDPDANILTA 72
|||||

QY 61 LNSPQGMGIRGADPQCFQQAAGLAGTFRATLSSRLQDLYSIVRRADRTGVPVNL 120
 Db 73 GLPFRPGVIRD-RGTQ---LQARAGEITAEW-AFI-----ADREGMAELV 114
 QY 121 RDVLEF-----PSWEALFSSGQKFGARI 146
 Db 115 RVEVALGRAVIANHHPHPIEPIWILGKAFKAVKANI 151

RESULT 5

GLI-CHICK
 ID - GLI-CHICK STANDARD; PRT; 556 AA.
 AC P55878;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein Gli1 (Gli) (Fragment).
 GN GLI1 OR GLI.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97105842; PubMed=8948590;
 RA Marigo V., Johnson R.L., Vortkamp A., Tabin C.J.;
 RT "Sonic hedgehog differentially regulates expression of GLI and Gli3
 during limb development.";
 RL Dev. Biol. 190:273-283 (1996).
 CC -!- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING
 CC NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT
 CC AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL
 CC NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE
 CC TRANSDUCTION OF SHH SIGNAL (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U60762; AAB51659.1; -.
 DR HSSP; P08151; 2GLI.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf_C2H2; 5.
 DR SMART; SM00355; Znf_C2H2; 5.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
 DR Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;
 KW Nuclear protein; Repeat.
 FT ZN_FING 247 272 C2H2-TYPE.
 FT ZN_FING 280 307 C2H2-TYPE.
 FT ZN_FING 313 337 C2H2-TYPE.
 FT ZN_FING 343 368 C2H2-TYPE.
 FT ZN_FING 374 399 C2H2-TYPE.
 FT ZN_FING 556 556 C2H2-TYPE.
 FT NON_TER 556
 SQ SEQUENCE 556 AA; 60215 MW; 722D2A5A1CA4D98 CRC64;

Query Match 7.3%; Score 90; DB 1; Length 556;
 Best Local Similarity 26.0%; Pred. No. 1.8;
 Matches 63; Conservative 22; Mismatches 93; Indels 64; Gaps 13;

QY 17 PQPYGAPH-----HGSVVHFQAPRTGPGPVHVT---THHQDFOLVHLVALNSPOP 66

Db 7 PQAPFAEHCCPPLHGAGAGTFLGLQGLDFPV-CHQPNLASHHGVLV-----PQTEHP 60

QY 67 GGMGIRGADPQCFQQAAGLAGTFRAP-----LSSRLQDLYSIVRRADRTGVPVNL 122

Db 61 GG-----AADSRSTPRGAGKLGKRALISPLSDSSVDLQTVRTSPNSLVAFINSRC 115
 QY 123 EVLPFSWEAL-----FSGSEGQLKFGARIFS-----FDCRDVLOHPA 159
 Db 116 ASAGGSYGHLSISITISPLGYQNPFGQKQGGQLFHTFPLPCSSHETLSRPGLIHPT 175
 QY 160 WPRKSVWHGSDPSGRR-----LTDSCYETWRTB-----APAATGQASSLLAGRL-----LE 205
 Db 176 PARGTIKHCQQLKRLSLSPFLTAKYPEE-KSEGDISSPASTGTQDPLL-GMLSVRDPLE 233
 QY 206 QE 207
 Db 234 KE 235

RESULT 6

NME3_HUMAN
 ID - NME3_HUMAN STANDARD; PRT; 1233 AA.
 AC Q14957;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Glutamate [NMDA] receptor subunit epsilon 3 precursor (N-methyl
 DE D-aspartate receptor subtype 2C) (NR2C) (NMDAR2C).
 GN GRIN2C.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97189248; PubMed=9037519;
 RA Lin Y.J., Boyetto S., Carver J.M., Giordano T.;
 RT "Cloning of the cDNA for the human NMDA receptor NR2C subunit and its
 RT expression in the central nervous system and periphery.";
 RL Brain Res. Mol. Brain Res. 43:57-64 (1996).
 CC -!- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels
 CC with high calcium permeability and voltage-dependent sensitivity
 CC to magnesium. Mediated by glycine.
 CC -!- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Mainly in brain with predominant expression in
 CC in the cerebellum, also present in the hippocampus, amygdala,
 CC caudate nucleus, corpus callosum, subthalamic nuclei and thalamus.
 CC Detected in the heart, skeletal muscle and pancreas.
 CC -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; L76224; AAA8096.1; -.
 DR HSSP; P19491; 1GR2.
 DR Genew; HGNC:4587; GRIN2C.
 DR MIM; 138254; -.
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0004972; P:N-methyl-D-aspartate selective glutamate re...; TAS.
 DR GO; GO:0007215; P:glutamate signaling pathway; TAS.
 DR GO; GO:0006810; P:transport; TAS.
 DR InterPro; IPR001828; ANF_receptor.
 DR InterPro; IPR001320; Ion_glu_receptor.
 DR InterPro; IPR001508; NMDA_receptor.
 DR InterPro; IPR001311; SRP_glu_receptor.
 DR Pfam; PF01094; ANF_receptor; 1.
 DR Pfam; PF00060; lig_chan; 1.
 DR PRINTS; PR00177; NMDARECEPTOR.
 DR SMART; SM00079; FBPE; 1.

KW Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
 KW Ionic channel; Magnesium.
 FT CHAIN 1 19
 FT CHAIN 20 1233
 FT TRANSMEM 554 574
 FT TRANSMEM 597 617
 FT TRANSMEM 627 647
 FT TRANSMEM 815 835
 FT SITE 612 612
 FT CARBOHYD 70 70
 FT CARBOHYD 337 337
 FT CARBOHYD 438 438
 FT CARBOHYD 539 539
 SQ SEQUENCE 1233 AA; 134239 MW; 700F4629EDD0699 CRC64;
 Query Match 7.1%; Score 88; DB 1; Length 1233;
 Best Local Similarity 26.6%; Pred. No. 6.8;
 Matches 63; Conservative 15; Mismatches 93; Indels 66; Gaps 12;
 Qy 9 AGPRLI---DPQYPCAPHGQVWFQPARPTGPGVHTHTHQDFQLVHLVALNSPQ 65
 Db 940 SGSPCLPTDPPEPSPTGWP-----PPGGR-----ALVRRAPQ 976
 Qy 66 PGGMRGIRGADFQCFQ-----QARAAGLAGTFRFLSSRLQDL-----YSTVREAD 111
 Db 977 PGGPPTPGPLSDVSRVSRPAREWVPVTCGCHGLSASERPLSPARCHYSSFPRAAD 1036
 Qy 112 RTGVVNVNLRDEVLPFQSWALFGSEGLQKPGARIFSGRVLQHPAPKRVWHGSDP 171
 Db 1037 RSGRFFL-----PLFPELDPLLGKEQL---AR-----REALHAAR-----GSRP 1077
 Qy 172 SGRLTDSYCTWR--TEAPA-ATGOASSLLAG-----RLLEQEAASCRHAFVVLGIE 221
 Db 1078 RHASLPVSVAEAFAPSPSLFAGCTGACARPDPGHSACRLAQASCLPIYREACQE 1134
 RESULT 7
 ATX1 HUMAN
 ID ATX1 HUMAN STANDARD; PRT; 816 AA.
 AC P54253; OSUJG2; Q9V4J1.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ataxin-1 (Spinocerebellar ataxia type 1 protein).
 DE SCAL OR ATX1.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Cerebellum;
 RX MEDLINE=95038838; PubMed=7951322;
 RA Banfi S., Servadio A., Chung M.-Y., Kwiatkowski T.J. Jr., McCall A.B.,
 RA Duvick L.A., Shen Y., Roth E.J., Orr H.T., Zoghbi H.Y.;
 RT "Identification and characterization of the gene causing type 1
 RT spinocerebellar ataxia.";
 RL Nat. Genet. 7:513-519(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tubby B.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 189-231 FROM N.A.
 RX MEDLINE=96177682; PubMed=8634720;
 RA Quan F., Janas J., Popovich B.W.;
 RT "A novel CAG repeat configuration in the SCAL gene: implications for
 RT the molecular diagnostics of spinocerebellar ataxia type 1.";
 RL Hum. Mol. Genet. 4:2411-2413(1995).
 RN [4]
 RP SELF-ASSOCIATION SITE.

RX MEDLINE=97252384; PubMed=9097953;
 RA Burridge E.N., Davidson J.D., Duvick L.A., Kosty B., Zoghbi H.Y.,
 RA Orr H.T.;
 RT "Identification of a self-association region within the SCAL gene
 RT product, ataxin-1.";
 RL Hum. Mol. Genet. 6:513-518(1997).
 RN [5]
 RP RNA-BINDING DOMAIN.
 RX MEDLINE=21065637; PubMed=11136710;
 RA Yue S., Serra H.G., Zoghbi H.Y., Orr H.T.;
 RT "The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding
 RT activity that is inversely affected by the length of its polyglutamine
 RL tract.";
 RL Hum. Mol. Genet. 10:25-30(2001).
 CC -1- FUNCTION: Binds RNA in vitro. May be involved in RNA metabolism.
 CC The expansion of the polyglutamine tract may alter this function.
 CC -1- SUBUNIT: Interacts with LAMP and ALU.
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=1;
 CC Comment=At least 2 isoforms are produced;
 CC Name=1;
 CC IsoId=P54253-1; Sequences=Displayed;
 CC -1- TISSUE SPECIFICITY: Widely expressed throughout the body.
 CC -1- POLYMORPHISM: The poly-gln region of SCAL is highly polymorphic (4
 CC to 39 repeats) in the normal population and is expanded to about
 CC 40-83 repeats in scal patients. Longer expansions result in
 CC earlier onset and more severe clinical manifestations of the
 CC disease.
 CC -1- DISEASE: Defects in SCAL are the cause of spinocerebellar ataxia
 CC type 1 (SCAL) [MIM:164400]; also known as olivopontocerebellar
 CC atrophy 1 (OPCA 1). SCAL is an autosomal dominant
 CC neurodegenerative disorder characterized by progressive neuronal
 CC loss in the cerebellum, brain stem and spinocerebellar tracts.
 CC Clinical features are cerebellar ataxia, dysarthria,
 CC ophthalmoparesis, muscle wasting and neuropathy. Onset of the
 CC disease usually occurs in the third or fourth decade of life and
 CC death occurs ten to twenty years later.
 CC -1- MISCELLANEOUS: The self-association seems to be necessary to form
 CC nuclear aggregates.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X79204; CAA55793.1; -;
 CC EMBL; AL009031; CAA15622.1; -;
 CC EMBL; S82497; AAD14401.1; -;
 CC FIR; S46268; S46268.
 CC Genew; HGNC:10548; SCAL.
 CC MIM; 601556; -;
 CC MIM; 164400; -;
 CC GO; GO:0005737; C:cytoplasm; TAS.
 CC GO; GO:0005634; C:nucleus; TAS.
 CC InterPro; IPR003652; Ataxin.
 CC SMART; SM00536; AXH; 1.
 CC RNA-binding; Nuclear protein; Polymorphism; Triplet repeat expansion;
 CC Alternative splicing.
 CC DOMAIN 197 226 POLY-GLN.
 CC DOMAIN 495 605 SELF-ASSOCIATION SITE.
 CC DOMAIN 541 767 RNA BINDING.
 CC DOMAIN 795 798 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY)
 SQ SEQUENCE 816 AA; 87051 MW; D49BA5DB423D0777 CRC64;
 Query Match 7.1%; Score 87.5; DB 1; Length 816;
 Best Local Similarity 22.4%; Pred. No. 4.7;
 Matches 54; Conservative 26; Mismatches 96; Indels 65; Gaps 10;

QY 8 LAGPRLDPPQYVGHGSGVYH-----QBARPTGGPVHTHTHODFOLVHLV 59
 Db 228 LSRAPGLITGSPPPA-QQNYVHSSPONTGRTASPPALPVHLHPHQ-----TWIPIHTL 282
 QY 60 ALNSPQGGWR-GIRGADFCQFQARAAGLAGTFRFLSSRLQD-----LVSIVRARAD 111
 Db 283 TLGPSPQVWQYADSGSHFVPREATKAE-----SSRLQQAIOAKEVLNGEMEKSR 333
 QY 112 RTGVVNVNLRDEVLFPSSWEALFGSGQLKPGARI--FSFGDRVLOHPAMPKSVVHGS 169
 Db 334 RYGA-----SSADJGLGKAGKGVHPYSHRVVHFS---PSDYSSR 374
 QY 170 DPSGRRLT-----DSYCETWRTAPATQASLLAGRLLEGEAASCSHAF 215
 Db 375 DPSGVASVMVNSNTPAADLEVOQAATHREASPSTLNDKSLGLHKGPHRSYALSPHTV 434
 QY 216 V 216
 Db 435 I 435

RESULT 8
 ALGB_PSEAE STANDARD; PRT; 494 AA.
 ID ALGB_PSEAE
 AC Q52453; Q9HY70;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Glycosyl transferase alg8 (EC 2.4.1.1-).
 GN ALG8 OR PA3541.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=88930;
 RX MEDLINE=94124011; PubMed=8294014;
 RA Maharaj R., May T.B., Wang S.-K., Chakrabarty A.M.;
 RT "Sequence of the alg8 and alg44 genes involved in the synthesis of
 RL alginate by Pseudomonas aeruginosa";
 RN Gene 136:267-269(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=2043737; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laibig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RL opportunistic pathogen.";
 RN Nature 406:959-964(2000).
 CC -!- FUNCTION: Possibly a processive enzyme that polymerizes GDP-
 CC mannuronic acid.
 CC -!- PATHWAY: Alginate biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L22611; AAC36875.2; ALT INIT.
 DR EMBL; AE004774; AAG06929.1; -;
 DR PIR; A83204; A83204.
 KW Alginate biosynthesis; Transferase; Glycosyltransferase;

KW Transmembrane; Complete proteome.
 FT TRANSMEM 12 34 POTENTIAL.
 FT TRANSMEM 49 71 POTENTIAL.
 FT TRANSMEM 381 403 POTENTIAL.
 FT TRANSMEM 423 445 POTENTIAL.
 FT CONFLICT 115 115
 SQ SEQUENCE 494 AA; 56456 MW; D89A5627E913EDF5 CRC64;

Query Match 7.08; Score 87; DB 1; Length 494;
 Best Local Similarity 24.68; Pred. No. 2.9; Mismatches 50; Gaps 7;
 Matches 45; Conservative 20;

QY 50 QDFOLVHLVALNSPQGGMRGIRGADP-----OCFOARAAGLAGTFRFLSSR 99
 Db 39 KDFILLICAVGIVWYRSMGGVFLRGLFLHVYVYFYRRVRQLGSAADPSHVFLMTSFR 98
 QY 100 LQDL-----YSIVRRADRTGVP-----VNLDRBVLFPSS-WEAL- 132
 Db 99 IDALTAMVTVRSVIREAIDSGYETTVCSIVSEMSDEVLSLWKNPDRVSLDFVRIIP 158
 QY 133 FSGSEGQLKPGARIFS-----FDGRDVLOHPAMPKSVVHGSDBS-GRRLTDS 179
 Db 159 GTGKRDGLAYGFRAISRHLFPDDDAVVAVDGDVLDGVVYKTVFVFKLFPNVGLLTINE 218
 QY 180 YCE 182
 Db 219 FCE 221

RESULT 9
 HGD_BRAJA STANDARD; PRT; 448 AA.
 ID HGD_BRAJA
 AC Q89XH1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Homogentisate 1,2-dioxygenase (EC 1.13.11.5) (Homogentisicase)
 DE (Homogentisate oxygenase) (Homogentisic acid oxidase).
 GN HMG4 OR BLU0343.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
 RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RL Bradyrhizobium japonicum USDA110.";
 RN DNA Res. 9:189-197(2002).
 CC -!- CATALYTIC ACTIVITY: Homogentisate + O(2) = 4-maleylacetate.
 CC -!- COFACTOR: Iron (By similarity).
 CC -!- PATHWAY: Catabolism of tyrosine; third step.
 CC -!- PATHWAY: Catabolism of phenylalanine; fourth step.
 CC -!- SIMILARITY: Belongs to the homogentisate dioxygenase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF005936; BAC45608.1; ALT INIT.
 DR HMANP; MF_00334; -;
 DR InterPro; IPR007113; Cupin_sup.
 DR InterPro; IPR005708; HmgA.
 Pfam; PF04209; HmgA; 1.

DR TIGRFAMS; TIGR01015; hmGA; 1.
KW Oxidoreductase; Dioxigenase; Metal-binding; Iron;
KW Phenylalanine catabolism; Tyrosine catabolism; Complete proteome.
FT METAL 346 346 IRON (BY SIMILARITY).
FT METAL 352 352 IRON (BY SIMILARITY).
FT METAL 382 382 IRON (BY SIMILARITY).
SQ SEQUENCE 448 AA; 49780 MW; F6DD465E68735D3C CRC64;

Query Match 6.9%; Score 86; DB 1; Length 448;
Best Local Similarity 24.0%; Pred. No. 3.2; Indels 36; Gaps 9;
Matches 44; Conservative 25; Mismatches 78;

QY 24 PH-----HGSVVHPQAPRTGGPVHTHTHODFQVLVHLVALNSPQGMGRGIRGA 75
DB 270 PHSIDVVAHGNVAPKYDLRTFSPVGAIGFDHPDPSIFTLTS-----PSETAGTANI 324

QY 76 DFQCFQQAAGLAGTPA-----FLSSRLQLYSIVRADRTGVP-VNLRDEVL--F 126
DB 325 DFVIFPE-RWVADNTFRPWTHTMNSPEMGLIYGVYDAKPGFVPGMSLHNCMLPHG 383

QY 127 PSWEALFSGSEGQKLP---GARIFDFGRDLVQHPAPKRSVWHGSDPSGRRLTDSYCE 182
DB 384 PDRDAFHASNGELKPKVLGTGVAFMFETR-----YFQRTVAHANAS--TLQDDYAD 434

QY 183 TWR 185
DB 435 CWK 437

RESULT 10
NRT1_CHICK
ID NRT1_CHICK STANDARD; PRT; 312 AA.
AC P55806;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NAD(P)+--arginine ADP-ribosyltransferase 1 precursor (EC 2.4.2.31)
OS (Mono(ADP-ribosyl)transferase 1) (Arl).
OC Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=White leghorn; TISSUE=Bone marrow;
RX MEDLINE=95050487; PubMed=7961658;
RA Tsuchiya M., Hara N., Yamada K., Osago H., Shimoyama M.;
RT "Cloning and expression of cDNA for arginine-specific ADP-
ribosyltransferase from chicken bone marrow cells."
RL J. Biol. Chem. 269:27451-27457(1994).
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N(2)-
(ADP-D-ribosyl)-L-arginine.
CC -!- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED
FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.
CC -!- SIMILARITY: Belongs to the Arg-specific ADP-ribosyltransferase
family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D31864; BAA06664.1; -
DR PIR; A55461; A55461.
DR InterPro; IPR000768; ART.
DR Pfam; PF01129; ART; 1.
DR PRINTS; PR00970; RIBTRNSFRASE.
DR PROSITE; PS01291; ART; 1.
KW Transferase; Glycosyltransferase; NAD; Signal; Zymogen.

FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 31
FT CHAIN 32 266
FT RIBOSYLTRANSFERASE 1.
FT PROPEP 267 312
FT ACT_SITE 231 231
FT ACT_SITE 231 231
SQ SEQUENCE 312 AA; 35318 MW; B82980439EC904FC CRC64;

Query Match 6.9%; Score 85; DB 1; Length 312;
Best Local Similarity 26.5%; Pred. No. 2.6;
Matches 53; Conservative 15; Mismatches 62; Indels 70; Gaps 13;

QY 32 FQPARTGPGVH--THYTHQ-----DFQLVHLVALNSPQ-----GMRGIRGADFCF 80
DB 118 FNEATRCQGRSHQDYTHSYHFKTLHFFLTQALFALRASQPCRYVYRGVGRIR-----F 171

QY 81 QQARAAGLAGTGERAFLSRL-----QDLYSVRRADRTGVEFVAL-----RDEVLF 126
DB 172 MTRQ--GKSVRFQGFSTSLKEATVNFQDITLVVVKTC--YGVFIKQFSPSPSEDEVLI 227

QY 127 PSWEAL-----FSGSEQ-----LKPGARIFSGFDGRDV-----LQH 157
DB 228 PPEVFEVINFSNDRGSKVIQLHSGKMSHTNCELLKPGQGMGRGHQEVGLGLSPGLSL 287

QY 158 PAMP-RKSVWHG-----SDP 171
DB 288 PVLPCRRRVWEGHREGDP 307

RESULT 11
CHRD_MOUSE
ID CHRD_MOUSE STANDARD; PRT; 948 AA.
AC Q9Z0E2; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chordin precursor.
GN CHRD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6SJL/F1;
RA Lu B., Bachiller D., Agius E., Piccolo S., De Robertis E.M.;
RT "BMP-binding domains in the chordin secreted protein."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99000848; PubMed=9782094;
RA Papano W.N., Scott I.C., Clark T.G., Eddy R.L., Shows T.B.,
RC Greenpan D.S.;
RT "Coding sequence and expression patterns of mouse chordin and mapping
of the cognate mouse chrd and human CHRD genes."
RL Genomics 52:236-239(1998).
CC -!- FUNCTION: Dorsalizing factor. Key developmental protein that
dorsalizes early vertebrate embryonic tissues by binding to
ventralizing TGF-beta family bone morphogenetic proteins (BMPs)
and sequestering them in latent complexes.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- DEVELOPMENTAL STAGE: Detected at high levels of a in 7 dpc mouse
embryos; its level decrease at later developmental stages and in
adult tissues.
CC -!- SIMILARITY: Belongs to the chordin family.
CC -!- SIMILARITY: Contains 4 CHRD domains.
CC -!- SIMILARITY: Contains 4 VMFC domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D31864; BAA06664.1; -
DR PIR; A55461; A55461.
DR InterPro; IPR000768; ART.
DR Pfam; PF01129; ART; 1.
DR PRINTS; PR00970; RIBTRNSFRASE.
DR PROSITE; PS01291; ART; 1.
KW Transferase; Glycosyltransferase; NAD; Signal; Zymogen.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF036276; AAD19895.1; -;
CC EMBL; AF069501; AAC68867.1; -;
CC MGI; MGI:1313268; ChrD.
CC GO; GO:0007420; P:brain development; IMP.
CC GO; GO:0007389; P:pattern specification; IMP.
CC GO; GO:0001501; P:skeletal development; IMP.
CC InterPro; IPR008559; SOG.
CC InterPro; IPR001007; VWF_C.
CC Pfam; PF00093; VWC; 4.
CC SMART; SM00566; SOG; 3.
CC SMART; SM00214; VWC; 4.
CC PROSITE; PS00933; CHR1; 4.
CC PROSITE; PS01208; WFC; 1; 2.
CC PROSITE; PS01184; WFC_2; 2.
CC Developmental protein; Repeat; Glycoprotein; Signal.
KW SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 948 CHORDIN.
FT DOMAIN 49 126 WFC 1.
FT DOMAIN 168 277 CHR1.
FT DOMAIN 279 398 CHR2.
FT DOMAIN 399 520 CHR3.
FT DOMAIN 526 646 CHR4.
FT DOMAIN 659 759 WFC 2.
FT DOMAIN 779 845 WFC 3.
FT DOMAIN 867 927 WFC 4.
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 877 877 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 948 AA; 101512 MW; 40C2DA01D9BD2147 CRC64;

Query Match 6.8; Score 84; DB 1; Length 948;
Best Local Similarity 24.2; Pred. No. 11;
Matches 53; Conservative 17; Mismatches 79; Indels 70; Gaps 9;
QY 15 LDPOAGLVEYRDPDEHSY---SDRGPGVGERTRADGHTDF-----VAL-----175
DB 132 LDPOAGLVEYRDPDEHSY---SDRGPGVGERTRADGHTDF-----VAL-----175
QY 69 MRGIRGADFCQQAARAGLAGTFRFLSSRLQDIYSIVRADRTGVPVNLREVLFP 128
DB 176 ---LTGPRQAVARARVLLSSRLSFSYQRLDRPSRVFTDPTG-----NILF--222
QY 129 WEALFSGEQLKPGARIFSPGDRVLQHPAPWRK-----SVHGSDPSGRLLTDCYET 183
DB 223 -----EHPATPTQDGLVCGVWRAVPELSVRLLR--EQ 253
QY 184 WRTEAPATQASLLAGRLLEQEAASCRHAPVVLCIEN 222
DB 254 LRVALVTST-HPSGEVWGPIWQALAEFTSAITLED 291

RESULT 12
ID HTS1 COCCA STANDARD; PRT; 5217 AA.
AC Q01886;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE HC-toxin synthetase (EC 6.3.2.-) (HTS).
GN HTS1.
OS Cochliobolus carbonum (Bipolaris zeicola).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=5017;
[1]
FP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 90305 / SB111;
RX MEDLINE=93100328; PubMed=1281482;
RA Scott-Craig J.S.; Panaccione D.G.; Pocard J.-A.; Walton J.D.;
RT "The cyclic peptide synthetase catalyzing HC-toxin production in the

filamentous fungus Cochliobolus carbonum is encoded by a
15.7-kilobase open reading frame.";
J. Biol. Chem. 267:26044-26049(1992).

[2]
FUNCTION.
STRAIN=ATCC 90305 / SB111;
MEDLINE=20138231; PubMed=10671527;
Cheng Y.-Q.; Walton J.D.;
"A eukaryotic alanine racemase gene involved in cyclic peptide
biosynthesis.";
J. Biol. Chem. 275:4906-4911(2000).
CC -!- FUNCTION: Non-ribosomal peptide synthetase, able to activate
proline and AEO (2-amino-9,10-epoxy-8-oxodecanoic acid), and
epimerize L-Pro. Catalyzes the production of HC-toxin; a cyclic
tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes
it to D-Pro; also uses D-Ala as a substrate but this is epimerized
from L-Ala by TOXG.
CC -!- COFACTOR: Contains 4 covalently bound phosphopantetheines.
CC -!- PATHWAY: HC-toxin non-ribosomal biosynthesis.
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC -!- SIMILARITY: Contains 4 acyl carrier domains.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-3 is the initiator.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M98024; AAA33023.1; -;
HSP; P14687; IAMU.
GO; GO:0016881; F:acid-D-amino acid ligase activity; IDA.
GO; GO:0009403; P:toxin biosynthesis; IDA.
InterPro; IPR000873; AMP-bind.
InterPro; IPR001242; Condensatn.
InterPro; IPR00163; Pp bind.
InterPro; IPR006162; Phantne S.
Pfam; PF00501; AMP-binding; 4.
Pfam; PF00668; Condensation; 5.
Pfam; PF00550; Pp-binding; 4.
PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
PROSITE; PS00455; AMP BINDING; 3.
PROSITE; PS00075; ACP_DOMAIN; 4.
Ligase, Multifunctional enzyme; Phosphopantetheine; Repeat.
KW REPEAT 249 842 DOMAIN 1.
FT REPEAT 1854 2452 DOMAIN 2.
FT REPEAT 3006 3606 DOMAIN 3.
FT REPEAT 4158 4738 DOMAIN 4.
FT DOMAIN 762 840 ACYL CARRIER (ACP) 1.
FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 2.
FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 3.
FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 4.
FT BINDING 803 803 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3568 3568 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4700 4700 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 5217 AA; 574637 MW; 0331D9C5400163A5 CRC64;

Query Match 6.7; Score 82.5; DB 1; Length 5217;
Best Local Similarity 24.9; Pred. No. 11e+02;
Matches 43; Conservative 15; Mismatches 72; Indels 43; Gaps 8;
QY 1 PWRADDILAGPRLDPPYGPAGPHGVSVMFPQAPRTGPGVHTHTHQDPQLVHLVA 60
DB 1713 PKRPTNVKAGEPE-----HHLCFEELSMSDPT-----EPDLFLV--1747
QY 61 LNSPQPGMGIRGADPOC--FOQARAGLAGTFRFLSSRLQDIYSIVRADRTGVPV 118
DB 1748 --ETQESNEVSAHLDFKLSIQAYATSIATVAHLSLVHDPY----RALNT-LRIV 1800

QY	1	PWRADDILAGPPRLDDQ	---PYP-----GAPHGSIYVHFP-----	-A 35
DB	131	PAGSDALLPGLPDLSPDLCGAPFEAFWEASPCAGAP	---SQCLYEPQLSPDPVKGLR 187	
QY	36	RPTGCGPHTHTHQDFQLVLHLVALNSPPCGMGRIGADFCQFQQAARAGLAGTTFRAF	95	
DB	188	RPASPALDAVSAPKGPYPAPWELLSVGADEGCSQG	---DYQAPEARF-----	PV 235
QY	96	LSSRLQDLYISVRADRTGVPVNVNLEVLFP	---WEALFGSGSQLKPGA 144	
DB	236	IGTKIEDLLSISCPAELPAVPA	-----NRLYPSGAYDA-FPLAPDGLGEA 280	

RESULT 14			
SR13	HUMAN	STANDARD;	PRT; 995 AA.
ID	SR13_HUMAN		
DT	28-FEB-2003	(Rel. 41, Created)	
DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	STAR-related lipid transfer protein 13 (StARD13) (START domain-containing protein 13) (46H23.2).		
DE	StARD13 OR Gt650.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI	TaxID=9606;		
[1]	SEQUENCE FROM N.A.		
RP	Rhodes S.;		
RA	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.		
RL	SEQUENCE FROM N.A.		
RA	Hunt A.;		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
CC	-!- FUNCTION: May function as a GTPase-activating protein.		
CC	-!- SIMILARITY: Contains 1 Rho-GAP domain.		
CC	-!- SIMILARITY: Contains 1 START domain.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcements or send an email to license@isb-sib.ch).		
CC	EMBL; AL049801; CAB42562.1; -		
DR	EMBL; Z84483; CAC94774.1; -		
DR	PIR; H59432; H59432		
DR	Genew; HGNC:19164; STARD13.		
DR	InterPro; IPR008936; RhoGAP.		
DR	InterPro; IPR000198; RhoGAP.		
DR	InterPro; IPR002913; START.		
DR	Pfam; PF006020; RhoGAP; 1.		
DR	Pfam; PF01852; START; 1.		
DR	SMART; SM00324; RhoGAP; 1.		
DR	SMART; SM00234; START; 1.		
DR	PROSITE; PS50238; RHO GAP; 1.		
DR	PROSITE; PS50848; START; 1.		
KW	GTPase activation.		
FT	DOMAIN 545 RHO-GAP.		
FT	DOMAIN 781 989 START.		
FT	SEQUENCE 995 AA; 111191 NW; 3F608FA94A8F8BF CRC64;		

Query Match	6.6%;	Score 82;	DB 1;	Length 995;
Best Local Similarity	23.0%;	Pred. No. 18;		
Matches	37;	Conservative	26;	Mismatches 52;
				Indels 46;
				Gaps 7;

QY	98	SELDQDLYISVRADRTGVP	-----VNVNLEVLFPSEWALFGSGSQLKPGA 143	
DB	36	SKVDDLYTLPRGDRNGSPGTCGRNNTSSSVLTDLSEPEVCSITHSSGSGDSRSRSPG	95	

QY 144 -----ARIFSPD-----CRDVLQHPANPRKSVHSGDPPSGRRRLTDSY-----CETWR 185
 Db 96 QCCTDNFVMDADFLVSSSLPQPPRDVLNHPFHPK-----NEKPRARAKSFLKRMETLR 149
 QY 186 TEAPATGQASS-----LLAGRLLEQEAASCRHAFVVLGIE 221
 Db 150 GKGAHGRHKGSGRTGLVISGPMQLQEPESFK---AMQCIIQ 187

RESULT 15

NR12_CHICK
 ID NR12_CHICK STANDARD; PRT; 312 AA.
 AC P55807;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NAD(P)(+)-arginine ADP-ribosyltransferase 2 precursor (EC 2.4.2.31)
 DE (Mono(ADP-ribosyl)transferase 2) (AT2).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=white leghorn; TISSUE=Bone marrow;
 RX MEDLINE=95050487; PubMed=7961658; Osago H., Shimoyama M.;
 RA Tsuchiya M., Hara N., Yamada K.,
 RT "Cloning and expression of cDNA for arginine-specific ADP-
 ribosyltransferase from chicken bone marrow cells.";
 RL J. Biol. Chem. 269:27451-27457(1994).
 CC -!- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N(2) -
 (ADP-D-ribosyl)-L-arginine.
 CC -!- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED
 CC FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.
 CC -!- SIMILARITY: Belongs to the Arg-specific ADP-ribosyltransferase
 family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; D31865; BAA06665.1; --
 DR PIR; B55461; B55461.
 DR InterPro; IPR000768; ART.
 DR Pfam; PF01139; ART.1
 DR PRINTS; PR00970; RIBTRNSFRASE.
 DR PROSITE; PS01291; ART; 1.
 KW Transferase; Glycosyltransferase; NAD; Signal; Zymogen.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 31 BY SIMILARITY.
 FT CHAIN 32 266 NAD(P)(+)-ARGININE ADP-
 FT PROPEP 267 312 RIBOSYLTRANSFERASE 2.
 FT ACT_SITE 231 231 POTENTIAL.
 FT SITE 231 231 BY SIMILARITY.
 SQ SEQUENCE 312 AA; 34966 MW; 9E45A9B7A6BC4057 CRC64;

Query Match 6.6%; Score 81.5; DB 1; Length 312;
 Best Local Similarity 26.5%; Pred. No. 5.3;
 Matches 53; Conservative 16; Mismatches 68; Indels 63; Gaps 13;

QY 22 GAPHHGSYVHFQPARTGPGVHTHTHQDFQFLVHLVALNSPQ-----GNGRGIRGA 75
 Db 125 GGSCHQYVHFY-----HFXT-LHFLLTQALFALRASQPCYVYVRGVRGIR-- 170
 QY 76 DFCQFOQARAAGLAGTFRFLSSRL-----ODLYSIVIRADRTGVPVNL-----R 121
 Db 171 -----FWTQR--GKSVRFQGTSTSLRKDVAVNFGQDTFFVWKTG--YGVPIKQSFYPSE 222

QY 122 DEVLFPSPWEAL-----FSGSEGQLKPGARI-----FSPGDRDVLQHPANPRKSVW-HGSDPSG 173
 Db 223 DEVLFPSPWEAL-----FSGSEGQLKPGARI-----FSPGDRDVLQHPANPRKSVW-HGSDPSG 173
 QY 174 RRLTD-----SYCETW 184
 Db 279 LGLSPGLALPVLPCNSGSCW 298

Search completed: March 26, 2004, 13:36:31
 Job time : 15.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:32:37 ; Search time 40.5556 Seconds
(without alignments)
1789.377 Million cell updates/sec

Title: US-09-938-391-2

Perfect score: 1239

Sequence: 1 PWRADDILAGPRLDPPQY.....CRHAFVVLCEIENSVMTSFSK 230

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mmc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_kvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1019	82.2	816	Q8N4S4	Q8N4S4 homo sapien
2	983.5	79.4	1140	Q61434	Q61434 mus musculus
3	897.5	72.4	226	Q9QZD2	Q9QZD2 rattus norv
4	837	67.6	261	Q8N4S4	Q8N4S4 homo sapien
5	835	67.4	187	Q8W419	Q8W419 homo sapien
6	792	63.9	1344	Q93419	Q93419 gallus gall
7	754	60.9	171	Q9WUW5	Q9WUW5 rattus norv
8	736.5	59.4	1315	Q8QHL9	Q8QHL9 xenopus lae
9	730	58.9	1307	Q8UFF7	Q8UFF7 xenopus lae
10	726	58.6	160	Q11QRT2	Q11QRT2 mus musculus
11	658.5	53.1	361	Q8AWC6	Q8AWC6 brachydanio
12	562.5	45.4	1367	Q35206	Q35206 mus musculus
13	561.5	45.3	1367	Q9EQD9	Q9EQD9 mus musculus
14	527	42.5	1388	Q9Y4W4	Q9Y4W4 homo sapien
15	457.5	36.9	950	Q86SC8	Q86SC8 ciona intes
16	456	36.8	102	Q96T70	Q96T70 homo sapien

ALIGNMENTS

RESULT 1

Q8N4S4

ID Q8N4S4

AC Q8N4S4

DT 01-OCT-2002 (TRENBLrel. 22, Created)

DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE Similar to collagen, type XVIII, alpha 1 (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC Tissue=Renal adenocarcinoma;

RA Strausberg R.;

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC033715; AAH33715.1; -.

DR InterPro; IPR008161; C1g_helix.

DR InterPro; IPR008160; Collagen.

DR Pfam; PF01391; Collagen; 5.

DR ProDom; PD000007; C1g_helix; 1.

KW Collagen.

FT NON_TER.

SQ SEQUENCE 816 AA; 82553 MW; 5D539B2946694F86 CRC64;

Query Match 82.2%; Score 1019; DB 4; Length 816;

Best Local Similarity 82.4%; Pred. No. 1.8e-89;

Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;

QY 1 PWRADDILAGPRLDPPQYPGAPHGSGVHFQAPRTGPGVHTHTHTHDFQLVHLVA 60

592 PWRADDILAGPRLDPPQYPGAPHGSGVHFQAPRTGPGVHTHTHTHDFQLVHLVA 647

61 LNSPQGMGIRGADQCFCQQAARAGLAGTFFAFLSSRLQDLYSIVRRADTGPVVNL 120

648 LNSPLSGMGRGIRGADQCFCQQAARAGLAGTFFAFLSSRLQDLYSIVRRADTGPVVNL 707

121 RDEVLFPSEALFSGSGQLKPGARIFSGDRVLQHPAPKRSVWVGSDPSGRRLTDSY 180

17 446.5 36.0 792 5 Q8MT89
18 434.5 35.1 299 5 Q8MSE3
19 428.5 34.6 778 5 Q8EBH1
20 369.5 29.8 650 5 Q17866
21 369.5 29.8 778 5 Q9U9K6
22 369.5 29.8 1117 5 Q9U9K7
23 369.5 29.8 1154 5 Q8I0G6
24 101.5 8.2 267 16 Q89I38
25 100.5 8.1 739 16 Q82BY9
26 100.5 8.1 745 16 Q8P380
27 99 8.0 244 2 Q53235
28 99 8.0 521 15 Q7ZL07
29 98 7.9 427 11 Q62700
30 98 7.9 521 15 Q7ZL05
31 98 7.9 2301 10 Q9VFA9
32 97.5 7.9 651 5 Q9VFA9
33 94.5 7.6 208 16 Q82K28
34 94 7.6 830 16 Q8CJX2
35 93.5 7.5 702 2 Q52660
36 93.5 7.5 713 2 Q52679
37 93 7.5 517 15 Q83378
38 91.5 7.4 1895 5 Q9VWS5
39 91 7.3 439 10 Q8LIQ7
40 91 7.3 6239 16 Q9S0R7
41 90 7.3 404 9 Q9XJK3
42 90 7.3 404 9 Q9XJK3
43 90 7.3 404 16 Q9KXC4
44 90 7.3 776 16 Q88E70
45 89 7.2 318 9 Q8SCB2

Q8nt89 drosophila
Q8mse3 drosophila
Q8ebh1 drosophila
Q17866 caenorhabdi
Q9u9k6 caenorhabdi
Q9u9k7 caenorhabdi
Q8i0g6 caenorhabdi
Q89i38 bradyrhizob
Q82by9 streptomyce
Q8p380 xanthomonas
Q53235 rhodobacter
Q7z107 recombinant
Q62700 rattus norv
Q72105 recombinant
Q9vfa9 drosophila
Q9vfa9 drosophila
Q82k28 rhizobium m
Q8cjx2 streptomyce
Q52660 escherichia
Q52679 escherichia
Q83378 rat leukemi
Q9vws5 drosophila
Q8liq7 oryza sativ
Q9s0r7 streptomyce
Q9xjk3 bacterioph
Q9xjk3 stx2 conver
Q9kxc4 escherichia
Q88e70 pseudomonas
Q8scb2 stx2 conver


```
Db 708 KDELFFSWALFSGSEGLPKGARIFSPDGRDLRHPPTWPKSVWHSDDNGRRLTESY 767
QY 181 CETWTEAPATQASLLAGLLEOEAASCRHAFVVLCIENSVMTS 227
Db 768 CETWTEAPATQASLLAGLLEOEAASCRHAFVVLCIENSVMTS 814

RESULT 2
Q61434 PRELIMINARY; PRT; 1140 AA.
AC Q61434;
DC 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muregaki Y., Yoshioaka H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region.";
RL Cell. Mol. Biol. Res. 196:576-582(1993).
DR EMBL; D17546; BAA04483.1; -
DR PIR; B56101; B56101.
DR HSSP; P39061; 1XOE.
DR MGD; MGI:88449; Col15a1.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PDC00007; Clg_helix; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1140 AA; 115156 MW; 880C7E6862B3BDFF CRC64;

Query Match 79.4%; Score 983.5; DB 11; Length 1140;
Best Local Similarity 80.1%; Pred. No. 7.4e-86;
Matches 185; Conservative 18; Mismatches 23; Indels 5; Gaps 2;

QY 1 PWRADDILAPRLDPQPPVGPAP-HGGSYVHFQPARPTGGPVHTHTHQQDFQLVHLV 59
Db 914 PWRADDILAPRLDPQPPVGPVPHHSSYVHLPPAPT-LSLAETHQDFQVLELV 969
QY 60 ALNSGPQGNRGIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPWN 119
Db 970 ALNTPLSGWMRGIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPWN 1029
QY 120 LRDEVLPSPSEALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDS 179
Db 1030 LRDEVLPSPSEALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDS 1089
QY 180 YCETWTEAPATQASLLAGLLEOEAASCRHAFVVLCIENSVMTSFSK 230
Db 1030 YCETWTEAPATQASLLAGLLEOEAASCRHAFVVLCIENSVMTSFSK 1140

RESULT 3
Q9QZD2 PRELIMINARY; PRT; 226 AA.
AC Q9QZD2;
DC 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Collagen XVIII (Fragment).
OS Rattus norvegicus (Rat).

Query Match 67.8%; Score 837; DB 4; Length 261;
Best Local Similarity 75.1%; Pred. No. 1.5e-72;
Matches 163; Conservative 16; Mismatches 24; Indels 14; Gaps 3;

QY 11 PPRLLDPQPPVGPAP-HGGSYVHFQPARPTGGPVHTHTHQQDFQLVHLV 70
Db 57 PPRLLDPQPPVGPAP-HGGSYVHFQPARPTGGPVHTHTHQQDFQLVHLV 102
QY 71 GIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPVNLDEVLFSWE 130
Db 103 GIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPVNLDEVLFSWE 162
QY 131 ALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDSYCYETWTEAPA 190
Db 163 ALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDSYCYETWTEAPA 222
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Perletti G., Concarini P., Giardini R., Marras E., Piccinini F.,
RA Folkman J., Chen L.;
RT "Antitumor activity of endostatin against carcinogen-induced rat
RT primary mammary tumors.";
RL Cancer Res. 50:1793-1796(2000).
DR EMBL; AF189709; AAF00975.1; -
DR HSRF; P39061; 1XOE.
FT NON TER
SQ SEQUENCE 226 AA; 25350 MW; 38B93C0486C0E949 CRC64;

Query Match 72.4%; Score 897.5; DB 11; Length 226;
Best Local Similarity 74.2%; Pred. No. 1.8e-78;
Matches 170; Conservative 18; Mismatches 28; Indels 13; Gaps 2;

QY 11 PPRLLDPQPPVGPAP-HGGSYVHFQPARPTGGPVHTHTHQQDFQLVHLV 61
Db 2 PPRLLDPQPPVGPVPHHSSHEHRRPPAPSPSP-ANTHQQDFQLVHLV 57
QY 62 NSPQPGMGRGIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPVNL 121
Db 58 NTPLSGWMRGIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPVNLK 117
QY 122 DEVLPSPSEALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDSYCY 181
Db 118 DEVLPSPSEALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDSYCY 177
QY 182 ETWTEAPATQASLLAGLLEOEAASCRHAFVVLCIENSVMTSFSK 230
Db 178 ETWTEAPATQASLLAGLLEOEAASCRHAFVVLCIENSVMTSFSK 226

RESULT 4
Q8NG19 PRELIMINARY; PRT; 261 AA.
AC Q8NG19;
DC 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Multi-functional protein MFP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Dou D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF282883; AAM52249.1; -
SQ SEQUENCE 261 AA; 26745 MW; CA60C920AF3E90E5 CRC64;

Query Match 67.8%; Score 837; DB 4; Length 261;
Best Local Similarity 75.1%; Pred. No. 1.5e-72;
Matches 163; Conservative 16; Mismatches 24; Indels 14; Gaps 3;

QY 11 PPRLLDPQPPVGPAP-HGGSYVHFQPARPTGGPVHTHTHQQDFQLVHLV 70
Db 57 PPRLLDPQPPVGPAP-HGGSYVHFQPARPTGGPVHTHTHQQDFQLVHLV 102
QY 71 GIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPVNLDEVLFSWE 130
Db 103 GIRGADFCQFQARAVGLSGTFRAFLSSRLQDLYSIVRRADRTGVPVNLDEVLFSWE 162
QY 131 ALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDSYCYETWTEAPA 190
Db 163 ALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVWHSDDNGRRLTDSYCYETWTEAPA 222
```

QY 191 ATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVMTS 227
 DB 223 ATGQASSLLGGRLLGQSAASCHHAYIVLICIENSVMTA 259

RESULT 5
 Q8WX15 PRELIMINARY; PRT; 187 AA.
 AC Q8WX15;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
 DE Collagen XVIII (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21409408; PubMed=11517600;
 RA Feng Y, Qui L.B., Liu C.X., Ma Q.J.;
 RT "Inhibition effect in vitro of purified endostatin expressed in Pichia
 pastoris";
 RL Sheng Wu Gong Cheng Xue Bao 17:278-282 (2001).
 RL EMBL; AF416592; AAL37720.1; -.
 FT NON TER 1
 SQ SEQUENCE 187 AA; 20448 MW; 72B1047D5838CD3 CRC64;

Query Match 67.4%; Score 835; DS 4; Length 187;
 Best Local Similarity 85.14; Pred. No. 1.6e-72;
 Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 47 HTHQDFQLVHLVALNSPQCGMEGIRGADFCQQAARAGLAGTFRFLSRLQDLYSI 106
 DB 5 HSHRDFQVHLVALNSPLSGMEGIRGADFCQQAARAGLAGTFRFLSRLQDLYSI 64

QY 107 VRRADRTGVVNVRLDEVLPFSWEALFSGEGQKFGARIFSDGRDVLQHPAPRKSVW 166
 DB 65 VRRADRAAAPTIVNLKDELLPFSWEALFSGEGQKFGARIFSDGRDVLQHPAPRKSVW 124

QY 167 HGSPSGRRLTDSYCTWRTTEAPATQAASSLLAGRLLEQEAASCRHAFVVLICIENSVMT 226
 DB 125 HGSDPNGRRLTESYCTWRTTEAPATQAASSLLAGRLLEQEAASCHHAYIVLICIENSVMT 184

QY 227 S 227
 DB 185 A 185

RESULT 6
 Q83419 PRELIMINARY; PRT; 1344 AA.
 AC Q83419;
 DT 01-NOV-1998 (TREMELrel. 08, Created)
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Collagen XVIII precursor.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98411346; PubMed=9738008;
 RA Halfter W, Dong S., Schurer B., Cole G.J.;
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan";
 RL J. Biol. Chem. 273:25404-25412 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX Halfter W, Dong S.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DB EMBL; AF083440; AAC33254.2; -.

HSP; P39061; IKOE.
 GO: GO:0005198; P: structural molecule activity; IEA.
 DR GO: GO:0007155; P: cell adhesion; IEA.
 DR InterPro; IPR008161; Cig_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; ConA_like_lectin.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 8.
 DR Pfam; PF02210; TSPN; 1.
 DR ProDom; PD000007; Cig_helix; 2.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 DR Collagen; Signal.
 KW COLLAGEN; SIGNAL.
 FT SIGNAL 1 26 POTENTIAL.
 SQ SEQUENCE 1344 AA; 137402 MW; 7AA36654FE940CCD CRC64;

Query Match 63.9%; Score 792; DB 13; Length 1344;
 Best Local Similarity 66.2%; Pred. No. 2.7e-67;
 Matches 155; Conservative 20; Mismatches 47; Indels 12; Gaps 4;

QY 1 PWRADIIAGPRLDPOYPGAPPH----GSYHFOFAPRTGPGVHTHTHODPOLVL 56
 DB 1119 PWRGDEVVAVQHHL--PQP-ELLQHELLNSYTHRWF-----DPAPVAHVHQDFQPAL 1170

QY 57 HLVALNSPQCGMEGIRGADFCQQAARAGLAGTFRFLSRLQDLYSIVRRADRTGVP 116
 DB 1171 HLVALNTPLSGMEGIRGADFCQQAARAGLAGTFRFLSRLQDLYSIVRRADRTAVP 1230

QY 117 VYNLDEVLPFSWEALFSGEGQKFGARIFSDGRDVLQHPAPRKSVWHDGSDGRRLL 176
 DB 1231 IVNLRDEVLPFSWEALFSGEGQKFGARILSFGDRDILQDSAWPQKSIWHGSDAKGRL 1290

QY 177 TDSYCTWRTTEAPATQAASSLLAGRLLEQEAASCRHAFVVLICIENSVMTSPSK 230
 DB 1291 PDSYCEAWRTDERTGTSQASSLSGKLEQSAASCHHAYIVLICIENSVMTAAK 1344

RESULT 7
 Q9WUW5 PRELIMINARY; PRT; 171 AA.
 AC Q9WUW5;
 DT 01-NOV-1999 (TREMELrel. 12, Created)
 DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE Collagen type XVIII, alpha (I) chain (Fragment).
 GN COL18A1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;
 RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;
 RT "Temporospatial expression of collagen XVIII/endostatin in acute and chronic liver injuries";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ236873; CAB44263.1; -.
 DR HSP; P39061; IKOE.
 FT NON TER 1
 FT NON TER 171 171
 SQ SEQUENCE 171 AA; 18933 MW; 81BE2E3FC2C9E72 CRC64;

Query Match 60.9%; Score 754; DB 11; Length 171;
 Best Local Similarity 81.9%; Pred. No. 9.1e-65;
 Matches 140; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

01-JUN-2003 (Tremblrel. 24, Last annotation update)
 Type XVIII collagen short variant.
 Xenopus laevis (African clawed frog).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=21216979; PubMed=12175494;
 Elam A., Peterson J., Pihlajaniemi T., Destree O.;
 "Cloning of three variants of type XVIII collagen and their expression
 patterns during Xenopus laevis development.";
 Mech. Dev. 114:109-113(2002).
 Mech. Dev. 114:109-113(2002).
 EMBL: AY052763; AAL14257.1;
 GO: GO:0005198; P: structural molecule activity; IEA.
 GO: GO:0007155; P: cell adhesion; IEA.
 InterPro: IPR008160; Collagen.
 InterPro: IPR003129; TSPN.
 Pfam: PF01391; Collagen; 6.
 Pfam: PF02210; TSPN; 1.
 SMART: SM00210; TSPN; 1.
 Collagen.
 SEQUENCE 1307 AA; 134153 MW; D53EDBFDE34976 CRC64;
 Query Match 58.9%; Score 730; DB 13; Length 1307;
 Best Local Similarity 63.0%; Pred. No. 2.5e-61;
 Matches 145; Conservative 24; Mismatches 49; Indels 12; Gaps 5;
 [1]
 QY 1 PWRADDILAGPRLDPOYPFGAPHGYSYVHFQPARPTGPGVHTHTTHDQFOLVHLVA 60
 DB 1090 PW-SDPRLPEPRVDPDPDPHGHGPHLEPS--PV---HTHQFNPALHLVA 1137
 QY 61 LNSPQGMGIRGADFCQFOQARAAGLACTFRAFLSRLQDLYSIVRRADRTGVPVNL 120
 DB 1138 LNAFLSGMKSIRGVDQCFEQARKAGLHTFRAFLSRLQDLYSIVRRADRTGVPVNL 1197
 QY 121 RDEVLPFSPWEALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVHSGDPSGRRITDSY 180
 DB 1198 RDEVLYDNWESLFSGSEAGMRFGARILSFQDKVATDPTWPKQVHSGDAKGRRLTESY 1257
 QY 181 CETWTEAPATGAQSSLLAGRLLEQEAASCRHAFVVLICIENSVMTSFSK 230
 DB 1258 CETWTEASVATGQASSLTSGKLEQRPQCNKFNIVLCIENSFMTNNRK 1307
 RESULT 10
 Q9CRT2 PRELIMINARY; PRT; 160 AA.
 ID Q9CRT2;
 AC Q9CRT2;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Procollagen, type XVIII, alpha 1 (fragment).
 GN COL18A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6J; TISSUE=Embryo;
 MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Kochiwa H.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Quackenbush J.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Wagner L., Washio T.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

55 VHLVALNSPQGMGIRGADFCQFOQARAAGLACTFRAFLSRLQDLYSIVRRADRTG 114
 DB 1 VHLVALNTPLSGMRGIRGADFCQFOQARAAGLACTFRAFLSRLQDLYSIVRRADRS 60
 QY 115 VPVNLDEVLPFSPWEALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVHSGDPSGR 174
 DB 61 VPIVNLKDEVLPSPWDVLFSGSQQLSGGARIISFGDRDVLQHPAPRKSVHSGDPSGR 120
 QY 175 RLTSYCYCTWTEAPATGAQSSLLAGRLLEQEAASCRHAFVVLICIENSVM 225
 DB 121 RLMSYCYCTWTEATGVTGQASSLLSGRLLEQEAASCHNSYIVLCIENSFM 171
 RESULT 8
 Q8QHL9 PRELIMINARY; PRT; 1315 AA.
 ID Q8QHL9;
 AC Q8QHL9;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Type XVIII collagen alpha chain.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 SEQUENCE FROM N.A.
 Ishino T., Sekimizu K., Natori S., Kubo T.;
 "Identification and characterization of genes expressed selectively in
 the regenerating tail of Xenopus laevis tadpole";
 Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 EMBL: AB047066; BAB84674.1;
 GO: GO:0005198; P: structural molecule activity; IEA.
 GO: GO:0007155; P: cell adhesion; IEA.
 InterPro: IPR008161; C1g helix.
 InterPro: IPR008160; Collagen.
 InterPro: IPR003129; TSPN.
 Pfam: PF01391; Collagen; 7.
 Pfam: PF02210; TSPN; 1.
 ProDom: PD000007; C1g helix; 1.
 SMART: SM00210; TSPN; 1.
 Collagen.
 SEQUENCE 1315 AA; 134946 MW; 0C56C235DE058365 CRC64;
 Query Match 59.4%; Score 736.5; DB 13; Length 1315;
 Best Local Similarity 62.6%; Pred. No. 6e-62;
 Matches 144; Conservative 24; Mismatches 55; Indels 7; Gaps 3;
 [1]
 QY 1 PWRADDILAGPRLDPOYPFGAPHGYSYVHFQPARPTGPGVHTHTTHDQFOLVHLVA 60
 DB 1093 PW-SDPRLPEPRVDPDPDPHGHGPHLEPS--PV---HTHQFNPALHLVA 1145
 QY 61 LNSPQGMGIRGADFCQFOQARAAGLACTFRAFLSRLQDLYSIVRRADRTGVPVNL 120
 DB 1146 LNAFLSGMKSIRGVDQCFEQARKAGLHTFRAFLSRLQDLYSIVRRADRTGVPVNL 1205
 QY 121 RDEVLPFSPWEALFSGSEGQKPGARIFSDGRDVLQHPAPRKSVHSGDPSGRRITDSY 180
 DB 1206 RDEVLYDNWESLFSGSEAGMRFGARILSFQDKVATDPTWPKQVHSGDAKGRRLTESY 1265
 QY 181 CETWTEAPATGAQSSLLAGRLLEQEAASCRHAFVVLICIENSVMTSFSK 230
 DB 1266 CETWTEASVATGQASSLTSGKLEQRPQCNKFNIVLCIENSFMTNNRK 1315
 RESULT 9
 Q8JFF7 PRELIMINARY; PRT; 1307 AA.
 ID Q8JFF7;
 AC Q8JFF7;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RL EMBL: AK014292; BAB29249.1; --
DR HSP: P39061; 1KOE.
DR MGD: MGI:89451; Coll8a1.
DR GO: GO:0005604; C:basement membrane; IDA.
DR GO: GO:0001525; P:angiogenesis; IMP.
FT NON_TER 1
SQ SEQUENCE 160 AA; 17725 MW; 60P853D777C375D2 CRC64;
Query Match 58.6%; Score 726; DB 11; Length 160;
Best Local Similarity 83.8%; Pred. No. 4.2e-62;
Matches 134; Conservative 16; Mismatches 10; Indels 0; Gaps 0;
QY 71 GIRGADFCQFQQAAGLAGTFRFLSSRLQDLYSIVRRADRTGVPVYNLRDEVLPFSWE 130
Db 1 GIRGADFCQFQQAAGLAGTFRFLSSRLQDLYSIVRRADRTGVPVYNLRDEVLPFSWD 60
QY 131 ALFSGSEGOLKPGARIFSGDGRVLOHPAPRKSVHGHSDPSGRRLTDSYCETWTEAPA 190
Db 61 SIFSGSGQLQPGARIFSGDGRVLOHPAPRKSVHGHSDPSGRRLTDSYCETWTEAPA 190
QY 191 ATQOASSLLAGRLDQEAASCHAFVVLCIENSVMTSFSK 230
Db 121 ATQOASSLLAGRLDQEAASCHAFVVLCIENSVMTSFSK 160
RESULT 11
QY QYAWC6 PRELIMINARY; PRT; 361 AA.
AC QYAWC6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Collagen XVIII (Fragment).
GN COL18A1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Haftek Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
RT "Sequence and embryonic expression of collagen XVIII NC11 domain
RT (endostatin) in the zebrafish.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ494837; CAD38825.1; --
FT NON_TER 1
SQ SEQUENCE 361 AA; 40222 MW; 3C5A0F8479D26735 CRC64;
Query Match 53.1%; Score 658.5; DB 13; Length 361;
Best Local Similarity 58.4%; Pred. No. 3.9e-55;
Matches 136; Conservative 23; Mismatches 55; Indels 19; Gaps 6;
QY 11 PP---RLDLP-QP---YPGAPHGHSVHFQPARPTGGPV-----RTHTHDQFQLVLH 57
Db 135 PPYTDNRVTDVENVIRYVQPEPNP---ITPARRPIPPVWPGGHATHHTSGP---GLH 188
QY 58 LVALNSPQGMGRGADGFCQQAAGLAGTFRFLSSRLQDLYSIVRRADRTGVPV 117
Db 189 LVALNSPQGMGRGADGFCQQAAGLAGTFRFLSSRLQDLYSIVRRADRTGVPV 117
QY 118 VNLKDEVLPFSWEALFSGSEGOLKPGARIFSGDGRVLOHPAPRKSVHGHSDPSGRRLT 177
Db 118 VNLKDEVLPFSWEALFSGSEGOLKPGARIFSGDGRVLOHPAPRKSVHGHSDPSGRRLT 177
Db 249 VNLKDEVLPFSWEALFSGSEGOLKPGARIFSGDGRVLOHPAPRKSVHGHSDPSGRRLT 177
QY 178 DSYCETWTEAPAATQOASSLLAGRLDQEAASCHAFVVLCIENSVMTSFSK 230
Db 309 DNYCETWRAGDRAVTLGLSSQLQGLLQOTSSSCSSSYIALCIENSVMTOSKK 361
RESULT 12
QY QY5206 PRELIMINARY; PRT; 1367 AA.
AC QY5206;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type XV collagen.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains.";
RL Genomics 45:31-41(1997).
DR EMBL: AF011450; AAC53387.1; --
DR PDB: 1DY2; 21-JAN-01.
DR MGD: MGI:88449; Coll15a1.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0007155; P:cell adhesion; IEA.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01391; Collagen; 5.
DR Pfam: PF02210; TSPN; 1.
DR ProDom: PD000007; Clg_helix; 1.
DR SMART: SM00282; LamG; 1.
DR SMART: SM00210; TSPN; 1.
DR Collagen.
KW Collagen.
SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3ABEC CRC64;
Query Match 45.4%; Score 562.5; DB 11; Length 1367;
Best Local Similarity 51.4%; Pred. No. 3.9e-45;
Matches 113; Conservative 26; Mismatches 60; Indels 21; Gaps 2;
QY 11 PPRLDLPQYPGAPHGHSVHFQPARPTGGPVTHTHDQFQLVLHVALNSPQGMGR 70
Db 1169 PPALSSNPYQPPPLNPILSANVERP-----VLHLVALNTPVAGDIR 1211
QY 71 GIRGADFCQQAAGLAGTFRFLSSRLQDLYSIVRRADRTGVPVYNLRDEVLPFSWE 130
Db 1212 ----ADFQCFQQAAGLAGTFRFLSSRLQDLYSIVRRADRTGVPVYNLRDEVLPFSWE 1267
QY 131 ALFSGSEGOLKPGARIFSGDGRVLOHPAPRKSVHGHSDPSGRRLTDSYCETWTEAPA 190
Db 1268 SIFSGSGQLQPGARIFSGDGRVLOHPAPRKSVHGHSDPSGRRLTDSYCETWTEAPA 190
QY 191 ATQOASSLLAGRLDQEAASCHAFVVLCIENSVMTSFSK 230
Db 1328 VTGPASPLSTGKILDKKAYSANRLIVLCIENSVMTDTRK 1367
RESULT 13
QY QY5209 PRELIMINARY; PRT; 1367 AA.
ID QY5209
AC QY5209;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Type XV collagen.
 GN COL15A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=97480713; PubMed=9339359;
 RA Hag P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
 RA Pihlajaniemi T.;
 RT "Cloning of mouse type XV collagen sequences and mapping of the
 RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
 RT (XV) collagen sequences indicates divergence in the number of small
 RT collagenous domains."
 RL Genomics 45:31-41(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=20522048; PubMed=11068203;
 RA Eklund L., Wuona A., Lietaud J., Pihlajaniemi T.;
 RT "Structure of the mouse type XV collagen gene, Col15a1, comparison
 RT with the human COL15A1 gene and functional analysis of the promoters
 RT of both genes."
 RL Matrix Biol. 19:489-500(2000).
 DR EMBL; AF261113; AAG27545.1; JOINED.
 DR EMBL; AF261109; AAG27545.1; JOINED.
 DR EMBL; AF261110; AAG27545.1; JOINED.
 DR EMBL; AF261111; AAG27545.1; JOINED.
 DR EMBL; AF261112; AAG27545.1; JOINED.
 DR EMBL; AF261113; AAG27545.1; JOINED.
 DR EMBL; AF261114; AAG27545.1; JOINED.
 DR EMBL; AF261115; AAG27545.1; JOINED.
 DR EMBL; AF261116; AAG27545.1; JOINED.
 DR EMBL; AF261117; AAG27545.1; JOINED.
 DR EMBL; AF261118; AAG27545.1; JOINED.
 DR EMBL; AF261119; AAG27545.1; JOINED.
 DR EMBL; AF261120; AAG27545.1; JOINED.
 DR EMBL; AF261121; AAG27545.1; JOINED.
 DR EMBL; AF261122; AAG27545.1; JOINED.
 DR EMBL; AF261123; AAG27545.1; JOINED.
 DR EMBL; AF261124; AAG27545.1; JOINED.
 DR EMBL; AF261125; AAG27545.1; JOINED.
 DR EMBL; AF261126; AAG27545.1; JOINED.
 DR EMBL; AF261127; AAG27545.1; JOINED.
 DR EMBL; AF261128; AAG27545.1; JOINED.
 DR EMBL; AF261129; AAG27545.1; JOINED.
 DR EMBL; AF261130; AAG27545.1; JOINED.
 DR HSP; P39061.1KOE.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; ConA like lec_gl.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 5.
 DR Pfam; PF02210; TSPN; 1.
 DR ProDom; PD000007; C1g_helix; 1.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 KW Collagen.
 SQ SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;

Query Match 45.3%; Score 561.5; DB 11; Length 1367;
 Best Local Similarity 51.4%; Pred. No. 4.8e-45;
 Matches 113; Conservative 26; Mismatches 60; Indels 21; Gaps 2;
 Cy 11 PRLDPPQYPGAPHGYSYVHFQAPRTGPGVTHTHQDQLVLHLVALNSPQGMGR 70

Db 1169 PFPALSNPQPPPLNPILSNYERP-----VHLVALNTPVAGDIR 1211
 QY 71 GIRGADFCQQAAGAGLCTAFATSSRLQDLYSIVRADRTGVPVNLDRDEVLPFSWE 130
 Db 1212 ---ADFQCFQQAAGLSTFAFLSSHLQDLSTVVRKAERFGLPIVNLKQVLNNWD 1267
 QY 131 ALFSGSEQQLKPGARIPSFQGRDVLQHPAPWRKSVHSGDPSGRRLTDSYCETWTEAPA 190
 Db 1268 SIFSGDGGQFNTHPIPIYSFQGRDVTDPSPWQKVVHGSNPYGVRLVDKYCEAWTTDMA 1327
 QY 191 ATQGAASSLLAGRLLEQBAASCRHAFVVLICIENSVMTSFSK 230
 Db 1328 VTGFASPLSTGKILDKQAYSANRLIVLCIENSFMVTDTRK 1367
 RESULT 14
 QY4W4
 ID QY4W4 PRELIMINARY; PRT; 1398 AA.
 AC QY4W4;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Type XV collagen.
 GN COL15A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 RT exon-intron organization in the 3' region of the corresponding gene."
 RL J. Biol. Chem. 269:4773-4779(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98316357; PubMed=9651385;
 RA Hag P.M., Wuona A., Lietaud J., Kivirikko S., Pihlajaniemi T.;
 RT "Complete exon-intron organization of the human gene for the alpha
 RT chain of type XV collagen (COL15A1) and comparison with the homologous
 RT COL18A1 gene."
 RL J. Biol. Chem. 273:17824-17831(1998).
 DR EMBL; L25280; AAC78500.1; JOINED.
 DR EMBL; AF052956; AAC78500.1; JOINED.
 DR EMBL; AF052957; AAC78500.1; JOINED.
 DR EMBL; AF052958; AAC78500.1; JOINED.
 DR EMBL; AF052959; AAC78500.1; JOINED.
 DR EMBL; AF052960; AAC78500.1; JOINED.
 DR EMBL; AF052961; AAC78500.1; JOINED.
 DR EMBL; AF052962; AAC78500.1; JOINED.
 DR EMBL; AF052963; AAC78500.1; JOINED.
 DR EMBL; AF052964; AAC78500.1; JOINED.
 DR EMBL; AF052965; AAC78500.1; JOINED.
 DR EMBL; AF052966; AAC78500.1; JOINED.
 DR EMBL; AF052967; AAC78500.1; JOINED.
 DR EMBL; AF052968; AAC78500.1; JOINED.
 DR EMBL; AF052969; AAC78500.1; JOINED.
 DR EMBL; AF052970; AAC78500.1; JOINED.
 DR EMBL; AF052971; AAC78500.1; JOINED.
 DR EMBL; AF052972; AAC78500.1; JOINED.
 DR EMBL; AF052973; AAC78500.1; JOINED.
 DR EMBL; AF052974; AAC78500.1; JOINED.
 DR EMBL; AF052975; AAC78500.1; JOINED.
 DR EMBL; L25285; AAC78500.1; JOINED.
 DR EMBL; L25284; AAC78500.1; JOINED.
 DR EMBL; L25283; AAC78500.1; JOINED.
 DR EMBL; L25282; AAC78500.1; JOINED.
 DR EMBL; L25281; AAC78500.1; JOINED.
 DR HSP; P39061.1KOE.
 DR GO; GO:0005198; F:structural molecule activity; IEA.

Job time : 45.5556 secs

DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like Lec-gl.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1388 AA; 141757 MW; 96828E45E847194B CRC64;

Query Match 42.5%; Score 527; DB 4; Length 1388;
Best Local Similarity 56.6%; Pred. No. 1e-41;
Matches 99; Conservative 27; Mismatches 45; Indels 4; Gaps 1;

QY 56 LHLVALNSPOGGMGIRGADFCQQAAGAGLAGTFRFLSSRLQDLYSIVRRADRTGV 115
Db 1218 LHLAALNMPVSGDIR---ADFQCFQQAAGAGLSTYRAFLSSRLQDLYSIVRKAERYSL 1273
QY 116 PVNLRLDEVLPFSWEALFSGSEQLKPGARIFSGDRDVLQHPAMPKRSVWHGSDPSGRR 175
Db 1274 PIVNLKGQVLFNWDSIFSGHGQFNWHPIYSFGDRDITDPSWPKVIMHGSSPHGVR 1333
QY 176 LTDSYCYETWTEAPATGQASSLLAGRLLEQAAASCRHAFVVLCTIENSVMTSFSK 230
Db 1334 LVNICYEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDARK 1388

RESULT 15

Q86SC8 PRELIMINARY; PRT; 950 AA.

AC Q86SC8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen XVIII homologue.
GN CI-COLLAGEN XVIII.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cloniidae; Ciona.
CX NCBI TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RA Hotta K., Takahashi H., Satoh N.;
RT "Further Characterization of Brachyury-Downstream Genes in Ciona
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076898; BAC57521.1; -
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001759; Pentaxin.
DR Pfam; PF01391; Collagen; 8.
DR PROSITE; PS00289; PENTAXIN; 1.
SQ SEQUENCE 950 AA; 94768 MW; 2767ECAL58CBA89F CRC64;

Query Match 36.9%; Score 457.5; DB 5; Length 950;
Best Local Similarity 55.2%; Pred. No. 3.3e-35;
Matches 96; Conservative 17; Mismatches 60; Indels 1; Gaps 1;

QY 53 QLVHLVALNSPOGGMGIRGADFCQQAAGAGLAGTFRFLSSRLQDLYSIVRRADR 112
Db 773 QKSLHMLALNPLRGNTGIVGADARCFQQAAGAGLSTYRAFLSSRLQDLYSIVRRADR 832
QY 113 TGVPVNLRLDEVLPFSWEALFSGSEQLKPGARIFSGDRDVLQHPAMPKRSVWHGSDPS 172
Db 833 RNVPVNLRLDEVLPFSWEALFSGSEQLKPGARIFSGDRDVLQHPAMPKRSVWHGSDPS 891
QY 173 GRLTDSYCYETWTEAPATGQASSLLAGRLLEQAAASCRHAFVVLCTIENSVM 226
Db 892 GRLNPMHYCASWYTDHKAVTGQASPLSTRLLAQKPYSCESGFFVLCVENSTRT 945

Search completed: March 26, 2004, 13:37:58

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 09:11:19 ; Search time 2233.23 Seconds
(without alignments)
10771.549 Million cell updates/sec

Title: US-09-938-391-3
Perfect score: 555
Sequence: 1 cacaccacaggattcca.....tgacctctttccaagtag 555

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.cm.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pin.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	555	100.0	555	6	AX399631	Sequence
2	555	100.0	555	6	BD188111	Method an
3	555	100.0	829	6	AX399629	Sequence
4	555	100.0	829	6	BD188110	Method an
5	550.4	99.2	552	6	BD240900	Expressio
6	419.6	75.6	2637	9	BC033715	Homo sapi
7	418	75.3	546	6	AR193165	Sequence
8	418	75.3	549	6	BD240882	Expressio
9	418	75.3	549	6	AX100086	Sequence
10	418	75.3	552	6	BD266741	Proteins
11	418	75.3	552	6	AX395662	Sequence
12	418	75.3	552	6	AX473835	Sequence
13	418	75.3	552	6	BD081407	Fused pro
14	418	75.3	563	6	AR431862	Sequence
15	418	75.3	786	9	AF282883	Homo sapi
16	418	75.3	900	6	AX370851	Sequence
17	418	75.3	1388	9	AKI30835	Homo sapi
18	418	75.3	2364	9	AK098216	Homo sapi
19	418	75.3	2376	6	AX816106	Sequence
20	418	75.3	2376	6	AX816108	Sequence
21	418	75.3	3394	6	IS1045	Sequence 4
22	418	75.3	3394	6	AX409531	Sequence
23	418	75.3	3394	9	HUMCOL18AX	Human colla
24	418	75.3	5408	9	AF018082	Homo sapi
25	418	75.3	5929	9	AF018081	Homo sapi
26	416.4	75.0	564	9	AF416592	Homo sapi
27	411.6	74.2	555	9	AF184060	Homo sapi
28	410.8	74.0	534	6	AR193166	Sequence
29	410.8	74.0	537	6	AX100092	Sequence
30	398.4	71.8	650	6	AX370853	Sequence
31	394.2	71.0	552	6	BD272245	Anti-angi
32	379	68.3	851	10	AF189709	Rattus no
33	375.8	67.7	555	6	BD102800	Process f
34	375.8	67.7	555	6	BD128313	Endostati
35	375.8	67.7	558	6	AR268609	Sequence
36	375.8	67.7	565	6	AR100842	Sequence
37	375.8	67.7	624	6	E34073	Carrier/DNA
38	375.8	67.7	624	6	AX084406	Sequence
39	375.8	67.7	1615	10	MMU03714	Mus muscu
40	375.8	67.7	4437	10	MUSCOLLAG	Mus muscu
41	375.8	67.7	4802	10	BC043697	Mus muscu
42	374.2	67.4	624	6	AX128409	Sequence
43	374.2	67.4	4031	6	IS1044	Sequence 1
44	374.2	67.4	4031	10	MUSCOLPRO	Mouse alpha
45	373.4	67.3	552	10	AF257775	Mus muscu

ALIGNMENTS

RESULT 1	AX399631	555 bp	DNA	linear	PAT 06-JUN-2002
AX399631	Sequence 3 from Patent EP1191036.				
LOCUS	AX399631				
DEFINITION	AX399631.1	GI:21335410			
ACCESSION	AX399631.1				
VERSION	AX399631.1				
KEYWORDS	Canis familiaris (dog)				
SOURCE	Canis familiaris				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
REFERENCE	1				
AUTHORS	Shepard, M.G. and Tong, S.				
TITLE	Methods and compositions for diagnosing and treating disorders involving angiogenesis				

Pred. No. is the number of results predicted by chance to have a

ACCESSION	BD198110
VERSION	BD198110.1 GI:32997849
KEYWORDS	JP 2003000268-A/1.
SOURCE	Canis sp.
ORGANISM	Canis sp.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
AUTHORS	Sheppard, M.G., and Tong, X.
TITLE	Method and composition for diagnosing and treating diseases relating to angiogenesis
JOURNAL	Patent: JP 2003000268-A 1 07-JAN-2003; PFIZER PRODUCTS INC
COMMENT	OS Canis sp. (dog) PN JP 2003000268-A/1 PD 07-JAN-2003 PF 24-AUG-2001 JP 2001254697 PR 25-AUG-2000 US 60727924 PI MICHAEL GEORGE SHEPPARD,XIAO TONG PC C12N15/09,A61K67/027,A61K31/7088,A61K35/76,A61K39/395 PC A61K39/395,A61K45/00, PC A61K48/00,A61P7/00,A61P9/00,A61P15/00,A61P17/00, PC A61P17/02, PC A61P17/06,A61P19/02,A61P19/08,A61P25/00,A61P27/02,A61P27/06, PC A61P29/00, PC A61P31/00,A61P35/00,A61P35/02,A61P35/04,A61P37/06,A61P43/00, PC C07K14/47, PC C07K16/18,C07K19/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC C12Q1/02,G01N33/15, PC G01N33/50,C12N15/00,C12N5/00 CC CANINE PRO ENDOSTATIN NUCLEOTIDE SEQUENCE FH Key Location/Qualifiers FT source 1..829 FT /organism='Canis sp. (dog)'. FT Location/Qualifiers 1..829 /organism='Canis sp.' /mol_type='genomic DNA' /db_xref='taxon:9616'
FEATURES	source 1..829 /organism='Canis sp.' /mol_type='genomic DNA' /db_xref='taxon:9616'
ORIGIN	Query Match 100.0%; Score 555; DB 6; Length 829; Best Local Similarity 100.0%; Pred. No. 1..1e-64; Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 CACACCCACGAGACTTCAGCTGGTGGCTGCACCTGGTGGCCCTGACAGCCGCGAGCCG 60
Db	139 CACACCCACGAGACTTCAGCTGGTGGCTGCACCTGGTGGCCCTGACAGCCGCGAGCCG 198
QY	61 GCGCGCATCGAGGCATCCGCGGAGCGAGCTTCAGTGTCTTCAGCAGGCGCGCGCGCGG 120
Db	199 GCGCGCATCGAGGCATCCGCGGAGCGAGCTTCAGTGTCTTCAGCAGGCGCGCGCGCGG 258
QY	121 GGGCTGGCGGCACCTTCGCGGGCCCTTCGTCGTCGGGGCTGCAGAGACCTCTACAGCATC 180
Db	259 GGGCTGGCGGCACCTTCGCGGGCCCTTCGTCGTCGGGGCTGCAGAGACCTCTACAGCATC 318
QY	181 GTGCGCGCGCGGACCGCACCGGGGGTCCCGTGTCTCAACTCTAGGACGAGTGTCTTTC 240
Db	319 GTGCGCGCGCGGACCGCACCGGGGGTCCCGTGTCTCAACTCTAGGACGAGTGTCTTTC 378
QY	241 CCCAGCTGGGAGGCTTATTCTCGGGGTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
Db	379 CCCAGCTGGGAGGCTTATTCTCGGGGTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 438
QY	301 TTCTCTTTTCGAGCGCAGGATGTCCTTCGACACCCCGCTGCGCCCGGAGACGCTGTGG 360
Db	439 TTCTCTTTTCGAGCGCAGGATGTCCTTCGACACCCCGCTGCGCCCGGAGACGCTGTGG 498
QY	361 CACGGCTCCGACCCACGCGGCGCGGCTTCACGACAGCTACTTCGAGACGTCGCGGACG 420
Db	499 CACGGCTCCGACCCACGCGGCGCGGCTTCACGACAGCTACTTCGAGACGTCGCGGACG 558

QY	301	TTCTCTTTTCGACGCGCAGAGATGCTCTGCAGCACC	CGCCCGCTGCCTGGCAGAGCGTG	GG 360
Db	301	TTCTCTTTTCGACGCGCAGAGATGCTCTGCAGCACC	CGCCCGCTGCCTGGCAGAGCGTG	GG 360
QY	361	CACGCGTCCGACGCCCGCAGCGCGCGCTGACCGC	AGCTACTGCGAGACGCTGGCGGACG	420
Db	361	CACGCGTCCGACGCCCGCAGCGCGCGCTGACCGC	AGCTACTGCGAGACGCTGGCGGACG	420
QY	421	GAGCGCCCGCGCGCCACCGCGCAGGCGTCTGCT	CGCTGCGCGGCGAGCGTCTGAGAGCAG	480
Db	421	GAGCGCCCGCGCGCCACCGCGCAGGCGTCTGCT	CGCTGCGCGGCGAGCGTCTGAGAGCAG	480
QY	481	GAGCGCGCGAGCTGCGCGCAGCGCTTCGCTGCT	CGCTGCTGCGATCGAGAACGCGTCATGACC	540
Db	481	GAGCGCGCGAGCTGCGCGCAGCGCTTCGCTGCT	CGCTGCTGCGATCGAGAACGCGTCATGACC	540
QY	541	TCCTTCTCCAAAG 552		
Db	541	TCCTTCTCCAAAG 552		
RESULT 6				
LOCUS	BC033715	2637 bp	linear	PRI 08-JUL-2002
DEFINITION	Homo sapiens, similar to collagen, type XVIII, alpha 1, clone IMAGE:4425380, mRNA, partial cds.			
ACCESSION	BC033715			
VERSION	BC033715.1 GI:21708148			
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 2637)			
TITLE	Strausberg, R.			
JOURNAL	Direct Submission			
REMARK	Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
COMMENT	NTH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapbs-k@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mc@nih.gov Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Griffith, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Staniripoo, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.I., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.			
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAC Plate: 68 Row: e Column: 16 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.			
source	Location/Qualifiers			
	1. .2637			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:4425380"			

[illegible]

LOCUS	AR193165	546 bp	DNA	linear	PAT 20-APR-2002	
DEFINITION	Sequence 4 from patent US 6346510.					
ACCESSION	AR193165					
VERSION	AR193165.1	GI:20239130				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 546)					
TITLE	O'Reilly,M.S. and Folkman,M.Judah.					
JOURNAL	Therapeutic antiangiogenic endostatin compositions					
FEATURES	Patent: US 6346510-A 4 12-FEB-2002;					
source	Location/Qualifiers					
	1..546					
	/organism="unknown"					
	/mol_type="unassigned DNA"					
ORIGIN						
	Query Match	75.3%	Score 418;	DB 6;	Length 546;	
	Best Local Similarity	85.3%;	Pred. No. 1.9e-46;			
	Matches 466;	Conservative	0;	Mismatches 80;	Indels 0;	Gaps 0;
QY	1	CACACCCACAGGACTTCCAGCTGTGCTGACACTGTGTGGCCCTGACAGCCGCGAGCCG	60			
DB	1	CACAGCACCCGCGACTTCCAGCCGCGTGTCCACCTGGTGTGGCTCAACAGGCCCTGTGTCA	60			
QY	61	GGCGCATGCGAGGATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCGCG	120			
DB	61	GGCGCATGCGGGGATCCCGGGGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCGCG	120			
QY	121	GGGTGGCGGACACTTTCGCGGCTTCTGTGCTGCGGCTGCAGGACCTCTACAGCATC	180			
DB	121	GGGTGGCGGACACTTTCGCGGCTTCTGTGCTGCGGCTGCAGGACCTGTACAGCATC	180			
QY	181	GTGGCGCGCGCCGACCGCAGCCGGGTGCGGTCGTCAACCTCAGGACAGAGTGCTCTTC	240			
DB	181	GTGGCGCGTGGCAGCCCGCAGCGGTGCCCATCTGTCAACCTCAAGACAGACTGCTTT	240			
QY	241	CCCAGCTGGGAGGCTTATTCTCGGCTCCGAGGCCCAGCTGAAGCCCGGCGCCGATC	300			
DB	241	CCCAGCTGGGAGGCTCTGTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGCGCACG	300			
QY	301	TTCTCTTTCAGCGCAGATGTCCTGACGACCCCGCTGGCCCCGGAAGAGCGTGTGG	360			
DB	301	TTCTCTTTCAGCGCAGGACGTCCTGAGGCAACCCCACTGGCCCCGGAAGAGCGTGTGG	360			
QY	361	CACGGCTCCGACCCCGCGCGCGCCCTGACCGACAGCTACTGCGAGACGTGCGGACG	420			
DB	361	CATGCTCGGACCCCAACGGCGCAGGCTGACCGAGAGCTACTGTGAGACGTGCGGACG	420			
QY	421	GAGCGCCCGGCGCACCGGGCAGCGCTGCTGCTGTGGCGGCGAGGCTGCTCGGACG	480			
DB	421	GAGGCTCCCTCGGCGCACGGGCCAGGCTCTCTGCTGCTGGGGGCGAGGCTCCTGGGCGAG	480			
QY	481	GAGCGCGGAGCTGCGCCACGCGCTTCGTGGTGTCTCTGCATCGAGAACAGCGTCAATGCC	540			
DB	481	AGTGGCGGAGCTGCCCATCAGCGCTACATCGTGTCTCTGCATTGAGACAGCTTCATGACT	540			
QY	541	TCCTTC	546			
DB	541	GCCTCC	546			

RESULT 8	BD240882	549 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD240882				
DEFINITION	BD240882				Expression and export of angiogenesis inhibitors as immunofusins.
ACCESSION	BD240882				
VERSION	BD240882.1	GI:33050652			
KEYWORDS	JP 2002523036-A/2.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				

RESULT 7
AR193165

and methods of use thereof.
BD266741
BD266741.1 GI:33076509
JP 2002532068-A/2.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Macdonald, N.J. and Sim, K.L.
Proteins that bind angiogenesis-inhibiting proteins, compositions
and methods of use thereof
Patent: JP 2002532068-A 2 02-OCT-2002;
ENTREMED INC
OS Homo sapiens (human)
EN JP 2002532068-A/2
PD 02-OCT-2002
PF 06-DEC-1999 JP 2000585271
PR 04-DEC-1998 US 09/206059
PI NICHOLAS J MACDONALD, KIM LEE SIM
PC C12N15/09, A61K38/00, A61P9/00, A61P27/02, A61P35/00, PC
A61P43/00,
PC
C07K14/705, G01N33/15, G01N33/50, G01N33/53, G01N33/566//A61K35/76, PC
A61K45/00,
PC A61K48/00, C12N15/00, A61K37/02
CC Proteins that bind angiogenesis-inhibiting proteins, CC
compositions and
CC methods of use thereof
FH Key Location/Qualifiers
FT source 1..552
FT /organism="Homo sapiens (human)"
FT Location/Qualifiers
1..552
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

Query Match 75.3%; Score 418; DB 6; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACGAGCTTCCAGCTGTGTGCTGACCTGTGGCTGACAGCCCGCAGCG 60
DB 1 CACAGCCACGCGACTTCCAGCCGCTGCTCCACCTGTGGCTCAACAGCCCTGTGCA 60
QY 61 GCGGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCG 120
DB 61 GCGGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCG 120
QY 121 GGGCTGCGCGGACCTTCCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 GGGCTGCGCGGACCTTCCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 GTGCG 240
DB 181 GTGCG 240
QY 241 CCAGCTGGAGGCGCTTATTCGCGCTCCGAGGCGCGCGCGCGCGCGCGCGCGCG 300
DB 241 CCAGCTGGAGGCGCTTATTCGCGCTCCGAGGCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 TTCTCTTTGAGCGGAGATGCTGCTGAGCAGCCCGCTGCGCGCGCGCGCGCGCGCG 360
DB 301 TTCTCTTTGAGCGGAGATGCTGCTGAGCAGCCCGCTGCGCGCGCGCGCGCGCGCG 360
QY 361 CACGCTCCGAGCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
DB 361 CATGGCTCGAGCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 421 GAGGCTCCG 480
DB 421 GAGGCTCCG 480

DB 421 GAGGCTCCG 480
QY 481 GAGGCTCCG 540
DB 481 AGTGCGCGGAGCTGCCATCAGCGCTACATCGTGTCTGCTGAGAGAGAGCTTCACTGACT 540
QY 541 TCCTTC 546
DB 541 GCCTCC 546

RESULT 11
AX395662
LOCUS AX395662
DEFINITION Sequence 30 from Patent WO0193897.
ACCESSION AX395662
VERSION AX395662.1 GI:21066471
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Sim, K.L. and Macdonald, N.J.
TITLE Angiotensin and endothelin binding proteins and methods of use
JOURNAL Patent: WO 0193897-A 30 13-DEC-2001;
EntreMed, Inc. (US)
FEATURES
source
1..552
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 75.3%; Score 418; DB 6; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACGAGCTTCCAGCTGTGTGCTGACCTGTGGCTGACAGCCCGCAGCG 60
DB 1 CACAGCCACGCGACTTCCAGCCGCTGCTCCACCTGTGGCTCAACAGCCCTGTGCA 60
QY 61 GCGGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCG 120
DB 61 GCGGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCG 120
QY 121 GGGCTGCGCGGACCTTCCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 GGGCTGCGCGGACCTTCCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 GTGCG 240
DB 181 GTGCG 240
QY 241 CCAGCTGGAGGCGCTTATTCGCGCTCCGAGGCGCGCGCGCGCGCGCGCGCGCG 300
DB 241 CCAGCTGGAGGCGCTTATTCGCGCTCCGAGGCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 TTCTCTTTGAGCGGAGATGCTGCTGAGCAGCCCGCTGCGCGCGCGCGCGCGCGCG 360
DB 301 TTCTCTTTGAGCGGAGATGCTGCTGAGCAGCCCGCTGCGCGCGCGCGCGCGCGCG 360
QY 361 CACGCTCCGAGCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
DB 361 CATGGCTCGAGCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 481 GAGGCTCCG 540
DB 481 AGTGCGCGGAGCTGCCATCAGCGCTACATCGTGTCTGCTGAGAGAGAGCTTCACTGACT 540

QY 541 TCCTTC 546
Db 541 GCCTCC 546

RESULT 12
AX473835
LOCUS AX473835 552 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 2 from Patent WO230982.
ACCESSION AX473835
KEYWORDS AX473835.1 GI:22208005
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Sim, K.L. and Liang, H.
TITLE Angiogenesis-inhibiting peptides and proteins and methods of use
JOURNAL Patent: WO 0230982-A 2 18-APR-2002;
JOURNAL Entremed, Inc. (US)

FEATURES
source Location/Qualifiers
1..552
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 75.3%; Score 418; DB 6; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACGAGACTTCAGCTGGTGTGCTGACCTGGTGGCCCTGAACAGCCGCGACCG 60
Db 1 CACAGCCACCGGACTTCAGCGCGTGTCTCCACCTGGTGGCTCAACAGCCGCGTCA 60

QY 61 GCGGCGATCGAGGACATCCGGGAGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 120
Db 61 GCGGCGATCGGGGCGATCCGGGGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 120

QY 121 GGGCTGCGCGGACCTTCGCGGCTTCCTGCTGCTGCGGCTGCGAGGACTTCACGATC 180
Db 121 GGGCTGCGCGGACCTTCGCGGCTTCCTGCTGCTGCGGCTGCGAGGACTTCACGATC 180

QY 181 GTGCGCGCGCGGACATCCGGGAGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 240
Db 181 GTGCGCGCGCGGACATCCGGGAGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 240

QY 241 CCGAGCTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGGCTGAGAGCGGCGCGGCG 300
Db 241 CCGAGCTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGGCTGAGAGCGGCGCGGCG 300

QY 301 TTCTCTTTTCAGCGGACAGATGCTCTCGAGCACCCCGCTGCGCCCGGAGAGCGTGG 360
Db 301 TTCTCTTTTCAGCGGACAGATGCTCTCGAGCACCCCGCTGCGCCCGGAGAGCGTGG 360

QY 541 TCCTTC 546
Db 541 GCCTCC 546

RESULT 13
BD081407
LOCUS BD081407 552 bp DNA linear PAT 27-AUG-2002
DEFINITION Fused protein containing angiotensin component and utilization thereof in antitumor therapy.
ACCESSION BD081407
VERSION BD081407.1 GI:22627010
KEYWORDS JP 2001518304-A/50.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Bolanowski, M.A., Caparon, M.H., Casperson, G.F., Gregory, S.A., Klein, B.K. and McKeown, J.P.
TITLE Fused protein containing angiotensin component and utilization thereof in antitumor therapy
JOURNAL Patent: JP 2001518304-A 50 16-OCT-2001;
JOURNAL GD SEARLE AND CO

COMMENT
OS Homo sapiens (human)
PN JP 2001518304-A/50
PD 16-OCT-2001
PF 30-SEP-1998 JP 20005113958
PR 01-OCT-1997 US 60/060609
PI MARK A BOLANOWSKI, MAIRE H CAPARON, GERALD F CASPERSON, SUSAN A GREGORY,
PI BARBARA K KLEIN, JOHN P MCKEOWN
PC C12N15/09, A61K38/00, A61K48/00, A61P9/10, A61P35/00, C07K14/52, PC C07K14/56,
PC C07K14/78, C12N9/68, C12N15/00, A61K37/02
CC Fused protein containing angiotensin component and utilization thereof in

FEATURES
source Location/Qualifiers
1..552
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 75.3%; Score 418; DB 6; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACGAGACTTCAGCTGGTGTGCTGACCTGGTGGCCCTGAACAGCCGCGACCG 60
Db 1 CACAGCCACCGGACTTCAGCGCGTGTCTCCACCTGGTGGCTCAACAGCCGCGTCA 60

QY 61 GCGGCGATCGAGGACATCCGGGAGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 120
Db 61 GCGGCGATCGGGGCGATCCGGGGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 120

QY 121 GGGCTGCGCGGACCTTCGCGGCTTCCTGCTGCTGCGGCTGCGAGGACTTCACGATC 180
Db 121 GGGCTGCGCGGACCTTCGCGGCTTCCTGCTGCTGCGGCTGCGAGGACTTCACGATC 180

QY 181 GTGCGCGCGCGGACATCCGGGAGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 240
Db 181 GTGCGCGCGCGGACATCCGGGAGCGGACTTCAGTGTCTTCAGAGCGGCGCGCGCG 240

QY 241 CCGAGCTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGGCTGAGAGCGGCGCGGCG 300
Db 241 CCGAGCTGGAGGCGCTTATCTCGGGCTCCGAGGCGGCGGCTGAGAGCGGCGCGGCG 300

QY 301 TTCTCTTTTCAGCGGACAGATGCTCTCGAGCACCCCGCTGCGCCCGGAGAGCGTGG 360
Db 301 TTCTCTTTTCAGCGGACAGATGCTCTCGAGCACCCCGCTGCGCCCGGAGAGCGTGG 360

361	CACGGCTCCGACCCACCGGCGCGCGCTGACCGACAGCTACTGCGAGACGCTGGCGGACG	420
	Qy	
361	CATGGCTCGGACCCCAACGGCGCGCAGGCTGACCGAGAGCTACTGTGAGACGCTGGCGGACG	420
	Db	
421	GAGCCCCGGGGGCGCACCGGGGCGAGGCGCTCGCTCGCTGGCGGGCAGAGCTGCTGGAGCAG	480
	Qy	
421	GAGGTCCTCTGGGCCACGGGCGCAGGCGCTCTCTGCTCTGGGGGGCAGGCTCTCTGGGGCGAG	480
	Db	
481	GAGGCGCGGAGCTGCCGCCACGCGCTTCGTGGTGCTCTGCATCGAGAACAGCGGTCTATGACC	540
	Qy	
481	AGTGGCGGAGCTGCCATCAGCGCTACATCGTGTCTCTGCATTGGAGAACAGCTTCATGACT	540
	Db	
541	TCCTTC	546
	Qy	
541	GCCTCC	546
	Db	

RESULT 14	AR431862	AR431862	Sequence 9 from patent US 6853098.	563 bp	DNA	linear	PAT 18-DEC-2003
	LOCUS	AR431862	AR431862				
	DEFINITION	AR431862	AR431862				
	ACCESSION	AR431862.1	GI:40194022				
	VERSION						

```

KEYWORDS
SOURCE
    .
    Unknown.
    Unknown.
    Unknown.
    Unknown.
    Unclassified.
    1 (bases 1 to 563)
    1 (bases 1 to 563)
REFERENCE
    Vieland, N. and Harding, E. I.
AUTHORS
    Method of producing mouse and human endostatin
TITLE
    Patent: US 6653098-A 9 25-NOV-2003;
JOURNAL
    Location/Qualifiers
FEATURES
    1..563
    source
        /organism="unknown"
        /mol_type="genomic DNA"

```

```

ORIGIN
Query Match 75.3%; Score 418; DB 6; Length 563;
Best Local Similarity 85.3%; Pred. NO. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```

QY

1 CACACCACGAGACTTCAGCTGGTGTGCACCTGGTGCGCCCTGAACAAGCAGCCGACGCG 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Ddb

6 CACAGCACCCGCACTCCAGCGGTGCCACCTGGTTGGCTCAACAGCCCCCTGTCA 65
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY

61 GGCGCATGGAGATCCGGGAGCGGAGATTCCAGTCTCCAGCAGGCGCGCGCGCG 120

Db	126	GAGGTGC	CGGGG	CGATCC	CGGGG	CCGACTTC	CGAGTCTT	CCAGCAGG	CGCGGG	CCGTG	125	
Qy	121	GAGGTGC	CGGGG	CGACCTTC	CGGGG	CCCTTC	GTGTCGT	CGGGG	CTTCAGG	CTTCACG	ATC	130
Db	126	GAGGTGC	CGGGG	CGACCTTC	CGGGG	CCCTTC	GTGTCGT	CGGGG	CTTCAGG	CTTCACG	ATC	135

[illegible]

	246	CCAGCTGGGAGAGCTGTGTTTCAGCGCTTGAGGTCGCTGAAGCCCGGGCAGCGATC	305
	301	TTCTCTTTCACGGCAGAGATGTCTCTGAGCACCCCGGCTGGCCCCCGAAGAGCGTGTGG	360

Accession	Sequence	Position
306	TTCTCTTTCAGCGCAAGGACGCTCTGAGGACACCCACCTTGGCCCCCAGAGAGAGCGTGTCG	385
361	CAGGCTCCGACCCACGCGGGGCCCTGTGACCGACGCTACTCTGAGAGCGTGGCGGACG	420
366	CATGCTCGACCCCAACGGGCGCAGGCTGACCCGAGAGCTACTCTGAGAGCGTGGCGGACG	425

421 GAGGCCCGGCGCCACCGGGCAGGCCTGTCGTGCTGGCGGCAAGCCTGCTGGAGCAG 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
426 GAGGCTCTCTCGGCCACGGGCCAGGCCTCTCGTGTGGGGGCAAGCCTCTGGGGCAG 485

100

Db 595 CATGGCTCGGACCCCAACGGGCGAGGCTGACCGAGAGCTACTGTGAGACGTGGCGGACG 654
Qy 421 GAGGCCCCGGGGCCACCGGCGAGGGCTGCTGCTGCTGGCGGCGAGGCTGCTGGAGCAG 480
Db 655 GAGGCTCCCTCGGCGACGGGCGAGGCTCCTGCTGCTGGGGGCGAGGCTCCTGGGGCAG 714
Qy 481 GAGGCGCGAGCTGCGGCGCACGCTTTCGTGGTGTCTGTCATCGAGAACAGCGTCAAGCC 540
Db 715 AGTGCGCGAGCTGCCATCAGCGCTACATCGTGTCTGCTGCTGAGAACAGCTTCATGACT 774
Qy 541 TCCTTC 546
Db 775 GCCTCC 780

Search completed: March 29, 2004, 13:38:51
Job time : 2235.23 secs

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 09:08:09 ; Search time 247.424 Seconds
(without alignments)
9529.183 Million cell updates/sec

Title: US-09-938-391-3

Perfect score: 555
Sequence: 1 cacaccaccaggacttcca.....tgactcttctccaagtag 555

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_29Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	555	100.0	555	6	AAL46063 Canine en
2	555	100.0	829	6	AAL46062 Canine pr
3	550.4	99.2	552	3	AAX51309 Canine an
4	419.6	75.6	1564	6	ABQ76740 DNA encod
5	418	75.3	546	3	AAX27004 Human end
6	418	75.3	549	2	AAX77719 Human end
7	418	75.3	549	3	AAX51291 Human ang
8	418	75.3	549	3	AAX29884 Human ang
9	418	75.3	549	5	AAS00867 Human gen
10	418	75.3	549	7	ABA00774 Human end
11	418	75.3	552	2	AAX35375 SEQ ID 50
12	418	75.3	552	3	AAX68203 Human end
13	418	75.3	552	3	AAC62023 Nucleotid
14	418	75.3	552	5	AAC88289 Human end
15	418	75.3	552	6	ABK50685 cDNA enco
16	418	75.3	555	4	AAL66529 Human vas
17	418	75.3	558	6	ABA99261 Human end
18	418	75.3	563	2	AAX08750 Human end
19	418	75.3	816	3	AAA64013 DNA encod
20	418	75.3	900	6	ABK09977 Synthetic
21	418	75.3	968	3	AAC62025 Nucleotid
22	418	75.3	3394	2	AAT94484 Human alp
23	418	75.3	3394	2	AAX78379 Human alp

24	418	75.3	3394	6	ABN95680 Gene #217
25	418	75.3	4551	6	ABN85301 Human col
26	418	75.3	4551	6	ABV94763 Human pan
27	418	75.3	4875	6	ABQ54955 Human ova
28	418	75.3	5408	2	AAX77720 Human col
29	417	75.1	551	6	ABQ81193 Human end
30	415.8	74.9	641	4	ABH79104 Human end
31	415.8	74.9	641	6	AAL55454 Specific
32	415.8	74.9	641	6	ABQ76079 Anticance
33	415.8	74.9	641	6	AAL44000 Reproduct
34	411.6	74.2	555	9	ADE40255 Human NOV
35	410.8	74.0	534	3	AAA27005 Alternate
36	410.8	74.0	537	5	AAS00868 Human gen
37	409.4	73.8	6462	3	AAA64014 Nucleotid
38	398.4	71.8	650	6	ABK09978 Synthetic
39	394.2	71.0	552	3	AAX50398 Human end
40	375.8	67.7	555	2	AAX79949 Endostati
41	375.8	67.7	555	2	AAX84635 Mouse en
42	375.8	67.7	555	5	AAC88290 Murine en
43	375.8	67.7	558	4	ABD18701 Mouse end
44	375.8	67.7	558	6	ABQ76173 Rat endos
45	375.8	67.7	558	6	ABQ76178 Rat endos

ALIGNMENTS

RESULT 1

AAL46063
ID AAL46063 standard; cDNA; 555 BP.

XX
AC AAL46063;

DT 19-JUL-2002 (first entry)

XX
DE Canine endostatin coding sequence.

XX
KW Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW rubecosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW plaque neovascularisation; telangiectasia; haemophilic joints;
KW angiofibroma; wound granulation; coronary collateral;
KW ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW cytotatic; antiarthritis; antipsoriatic; antidiabetic; ophthalmological;
KW gynaecological; gene; ss.

XX
OS Canis familiaris.

XX
FH Key Location/Qualifiers
FT CDS 1..555
FT /tag= a
FT /product= "endostatin"
FT /partial
FT /note= "no start codon"

XX
EP1191036-A2.

XX
27-MAR-2002.

XX
24-AUG-2001; 2001EP-00307224.

XX
25-AUG-2000; 2000US-0227924P.

XX
(PFIZ) PFIZER PROD INC.

XX
Sheppard MG, Tong X;

XX
WPI; 2002-354068/39.

XX
P-PSDB; AAO17430.

XX
PT An isolated nucleic acid molecule for the treatment of angiogenesis-

319	Db		GTGCGCGCGCGACCGAGGTCGCCGTCAACTCAGGACAGAGTGTCTTC	378
241	Qy	CCCAGCTGGGAGGCTTATTCTCGGGTCCGAGGGCCAGCTGAAGCCCGGGGCCCGGCATC	300	
379	Db	CCCAGCTGGGAGGCTTATTCTCGGGTCCGAGGGCCAGCTGAAGCCCGGGGCCCGGCATC	438	
301	Qy	TTCTCTTTTCGACGGCAGAGATGTCTCTCAGCACCCCGCTGGCCCCCGGAAGAGCGTGTGG	360	
439	Db	TTCTCTTTTCGACGGCAGAGATGTCTCTCAGCACCCCGCTGGCCCCCGGAAGAGCGTGTGG	498	
361	Qy	CACGGCTCCGACCCACGCGGGCGCGCTCACCACAGCTACTTCGACAGACGTGGCGGACG	420	
499	Db	CACGGCTCCGACCCACGCGGGCGCGCTCACCACAGCTACTTCGACAGACGTGGCGGACG	558	
421	Qy	GAGGCCCGGGCGGCACCGGGCAGGCGTCTGCTGCTGGCGGGCAGGCTGCTGAGAGAG	480	
559	Db	GAGGCCCGGGCGGCACCGGGCAGGCGTCTGCTGCTGGCGGGCAGGCTGCTGAGAGAG	618	
481	Qy	GAGGCCCGGAGCTCCGCGCACCGCTTCGTGGTGTCTTGTCATCGAGAACAGCGTCATGACC	540	
619	Db	GAGGCCCGGAGCTCCGCGCACCGCTTCGTGGTGTCTTGTCATCGAGAACAGCGTCATGACC	678	
541	Qy	TCCTTCTCCAAAGTAG	555	
679	Db	TCCTTCTCCAAAGTAG	693	

RESULT 3

RESUL 3
AAZ51309
ID AAZ51309 standard; DNA: 552 BP.

AA
AC
AA751309:

AA	06-JUN-2000 (first entry)
DT	

XX
DE
Canine angiogenesis inhibitor; endostatin DNA.

Canine; immunoglobulin Fc fragment; endostatin; immunofusin;
angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
vasotropic; vulnery; treatment; antiarteriosclerosis; tumour;
metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
ocular angiogenic disease; diabetic retinopathy; macular degeneration;
myocardial angiogenesis; plaque neovascularisation; telangiectasia;
wound granulation; keloid scar; gene therapy; ds.

XX
QS
Canis familiaris

XX	Key	Location/Qualifiers
FF	CDS	1..552
FT		/*tag= a
FT		/product= "Endostatin"
FT		/note= "Does not include stop codon"
FT		/partial

AA
PN
WO2000011033-A2XX
PD
02-MAR-2000

XX
25-AUG-1999
99W0-IIS019329

XX
DB 25-AUG-1999. 0911C-0007993D

XX
DA /1EVT-1 LEVYINGEN RHARM CORP

XX
PT LO K TI Y Gillies SD:

XX
DB WPT. 3000-237516/30

DR P-PSDB; AAY70265.

Novel fusion protein of angiotensin or endostatin and an immunoglobulin PT
PT PC region, useful for treating conditions mediated by angiogenesis, such XX

as rheumatoid arthritis, tumors and macular degeneration.

Example 8: Page 58-59: 68pp: English:

The patent discloses a DNA molecule encoding a fusion protein comprising a signal sequence, an immunoglobulin Fc region, and an angiogenesis inhibitor selected from angiostatin, endostatin, and a plasminogen fragment having angiostatin activity, a collagen XVII fragment having endostatin activity, or combinations of them. The fusion protein (immunofusin) is used to inhibit angiogenesis and to treat diseases or conditions mediated by angiogenesis. Conditions that may be treated include solid tumours, blood born tumours, tumour metastasis, benign tumours including haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases (e.g. diabetic retinopathy, retinopathy of prematurity, macular degeneration, corneal graft rejection, neovascular glaucoma, retrolental fibroplasia, rubeosis and Osher-Weber syndrome, myocardial angiogenesis, plaque neovascularisation, telangiectasia, haemophilic joints, angioma, angioma, wound granulation, and excessive or abnormal stimulation of angioblasts, intestinal cells, atherosclerosis, sclerodermal and endothelial cells, hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in gene therapy. The present sequence is a DNA encoding canine endostatin gene fragment used in the construction of immunofusin containing canine immunoglobulin Fc fragment

Sequence 552 BP: 77 A; 204 C; 190 G; 81 T; 0 U; 0 Other;

99.2%: Score 550.4: DB 3: Length 552:

Best Local Similarity 99.8%; Pred. No. 1e-82;

	Conservative	Mismatches	Indels	Gaps
	551;	0;	1;	0;

1 CACACCCACGAGCTTCAGCTGGTGTGCACCTGGTGGCCCTGAACAGCCGAGCCG 60

1 CACACCCACGAGCTTCAGCCGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG 60

61 GCGGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCGG 120

61 GCGGGCATGCGAGGCA TCCGGGAGCGGACTTCCAGTGCTTCCAGCGCGCGCGCG 120

121 CCGCTGGCCGGCAGCTTCGGGCCCTCTGTCGCGGCTGCAGGACCTCTACAGCATC 180

121 CCGCTGGCCGGCAGCTTCCTGTCGTGGGGCTGCAGGACCTCTACAGCATC 180

191 GTTCGGCGGCGGCGGCACCGGGTGGCCGTCAACCTCAGGACGAGGTGATCTTC 240

[illegible]

CCTCCGCGCAGTGGTGATTTTCACCAATTGCCTGTG

309

311 300

360

360

[illegible]

420

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

541 TCCTTCTCCAAG 552

541 TCCTTCTCCAAG 552

RESULT 4

PS Claim 10; Page 39; 69pp; English.

CC The present sequence encodes an endostatin protein which is the carboxy
CC terminal protein of human collagen XVIII. Recombinant mouse endostatin
CC (20 mg/kg) was administered subcutaneously to mice implanted with Lewis
CC lung carcinomas. There was tumour mass regression non-detectable levels
CC after 12 days of therapy due to the angiogenesis inhibitory activity of
CC endostatin. Thus the protein is useful for treatment of angiogenesis-
CC dependent cancers. The polynucleotide and polypeptide sequences of this
CC endostatin are useful for treating and diagnosis of tumours, ocular
CC angiogenic diseases, Osler-Webber syndrome, myocardial angiogenesis,
CC plaque neovascularisation, telangiectasia, haemophilic joints,
CC angiofibroma and wound granulation, for treatment of diseases related to
CC excessive or abnormal stimulation of endothelial cells e.g. intestinal
CC adhesions, atherosclerosis, scleroderma. The protein may also be useful
CC as a birth control agent by reducing or preventing uterine
CC vascularisation. The gene for endostatin may be isolated from cells or
CC tissue that express high levels of endostatin, eg. tumour cells, by
CC generating cDNA from mRNA using reverse transcriptase and then amplifying
CC the DNA sequence

XX

Sequence 546 BP; 80 A; 196 C; 177 G; 93 T; 0 U; 0 Other;

Query Match	75.3%;	Score 418;	DB 3;	Length 546;
Best Local Similarity	85.3%;	Pred. No. 8.9e-61;		
Matches 466;	Conservative 0;	Mismatches 80;	Indels 0;	Gaps 0

QY	1	CACACCCACAGAGCTTCACGTGGTGTGCACCTGTGGCCCTGAACAGCCGCGAGCGG	60
Db	1	CACAGCACCGGAGCTTCACGCGGTGTCTCCACCTGTGTTCGCTCAACAGCCCCCTGTCA	60
QY	61	GGCGGCATCGAGGAGATCGGGGAGCGGACCTCCAGTGTCTTCAGACGAGCGCGCCCGG	120
Db	61	GGCGGCATCGGGGGCATCGCGGGGCCGACTTCCAGTGTCTTCAGACGAGCGCGGGCCGTG	120
QY	121	GGGTGGCGCGGACCTTCGGGCGCTTCTGTGTGTGGCGGTGCAGGACCTCTACAGATC	180
Db	121	GGGTGGCGCGGACCTTCGGGCGCTTCTGTGTGTTCGCGCTCGAGACCTGTACAGCATC	180
QY	181	GTGCGCGCGCGCCGACCGCACCGGGGTGCCGTGTCAACTCAGGAGCAGAGTGCTCTTC	240
Db	181	GTGCGCGCGTGCACCGCGGCGAGCGTGCCCATCGTCAACTCAAGAGCAGAGTGCTGT	240
QY	241	CCGAGTGGAGAGCCCTTATTCGGGTCCGAGGGCCAGCTGAAGCCCGGGCCCGCATC	300
Db	241	CCGAGTGGAGAGCTCTGTTCAGGGTCTGAGGGTCCGCTGAAGCCCGGGCAGCGCATC	300
QY	301	TTCTCTTTTCGACGGCAGAGATGCTCCTCAGCACCCCGCCCTGGGCCCGGAAGAGCTGTGG	360
Db	301	TTCTCTTTGACGGCAGAGAGCTCTCTAGAGCACTCCAGCTGGGCCCCAGAGAGCTGTGG	360
QY	361	CACGGCTCGACCCGAGGGCGCGCTGACCGACAGCTACTGTGAGACGTGTGGCGGACG	420
Db	361	CATGCTCGGACCCCAAGGGCGGAGCTGACCGAGAGCTACTGTGAGACGTGTGGCGGACG	420
QY	421	GAGGCCCGGGGGGCACCGGGGAGGCGCTCGTGTGCTGTGGCGGACAGGCTGTGTGAGCAG	480
Db	421	GAGGCTCCCTCGGCCACGGGCGAGGCTCTCTGCTGTGTGGGGGCGAGCTCTGTGGGCGAG	480
QY	481	GAGGCGGAGAGCTGGCGGCACGCTTGTGGTGTCTGTGCATCGAGAACAGCGTCAAGCC	540
Db	481	AGTGGCGGAGCTGCCATCAGCGCTACATCGTCTCTGCATTGAGAACAGCTTCAATGACT	540
QY	541	TCCTTC 546	
Db	541	GCCTCC 546	

RESULT 6
AAAX77719
ID AAX77719 standard; DNA; 549 BP.
XX
AAAX77719;

XX	10-AUG-1999	(first entry)
DT		
XX		
DE	Human endostatin DNA coding region fragment.	
XX		
KW	Plasminogen; human; angiotatin; endostatin; gene therapy; vector;	
KW	anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;	
KW	tumour growth; solid tumour; diabetic retinopathy; retina; ss.	
XX		
OS	Homo sapiens.	
OS		
PN	WO9926480-A1.	
XX		
PD	03-JUN-1999.	
XX		
PF	20-NOV-1998; 98WO-US024950.	
XX		
PR	20-NOV-1997; 97US-00975424.	
XX		
PA	(GENE-) GENETIX PHARM INC.	
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.	
XX		
PI	Leboulch P, Pawliuk RJ, Bachelot T;	
XX		
DR	WPI; 1999-357696/30.	
DR	P-PSDB; AAY08693.	
XX		
PT	Anti-angiogenic gene therapy vectors.	
XX		
PS	Disclosure; Page 74; 83pp; English.	
XX		
CC	This invention describes a novel viral gene therapy vector comprising a	
CC	nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from	
CC	human or murine angiotatin, human or murine endostatin and angiogenesis-	
CC	inhibiting fusions and fragments, where the viral vector is sufficiently	
CC	attenuated for use in human gene therapy. The products of the invention	
CC	have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological	
CC	activity. The vector is used in gene therapy for inhibiting tumour growth	
CC	in humans harbouring a solid tumour. The vector expresses an anti-	
CC	angiogenic polypeptide. An additional use comprises treatment of diabetic	
CC	retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis	
CC	in the vicinity of the retina. The vector is administered to cells ex	
CC	vivo and then administered to the patient	
XX		
SQ	Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;	
	Query Match	75.3%; Score 418; DB 2; Length 549;
	Best Local Similarity	85.3%; Pred. No. 8,9e-61;
	Matches 466; Conservative	0; Mismatches 80; Indels 0; Gaps 0
QY	1	CACACCCACGAGGACTTCCAGCTGGTGTGTGACCTGGTGGCCCTGAACAGCCCGCAGCGG 60
DB	1	CACAGCCACCGCGACTTCCAGCGCGTGTCCACCTGGTTCGCTCAACAGCCCCCTGTCA 60
QY	61	GGCGGCATGCGAGGATCTCGGGGAGCGGACTTCAGTGTCTCCAGCAGCGCGCGCCCGG 120
DB	61	GGCGGCATGCGGGGCATCCGCGGGGCCGACTTCCAGTGTCTTCACGACGGCGGCGCGTG 120
QY	121	GGGTGTGGCGGCACCTTCCGGGCGCTTCCTGTGTGTGTGGCTGCAGGACCTCTACAGCATC 180
DB	121	GGGTGTGGCGGCACCTTTCGCGCGCTTCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
QY	181	GTGGCGCGCGCGACCCGACCGGGGTGCCCGTGTCTCAACTCAGGGGACGAGTGTCTTTC 240
DB	181	GTGGCGCGGTGCGGACCGCGCAGCGGTGCCCATCTGTCAACTCAGGACGAGTGTCTGT 240
QY	241	CCGAGCTGGGAGGCGCTATTCTCGGGCTCGAGGGCCAGCTCAAGCCGGGCGCGCATC 300
DB	241	CCGAGCTGGGAGGCGCTGTGTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGGACGCGATC 300
QY	301	TTCTCTTTTCGACGCGAGAGATGTCTCTGACGACCCCGCTGTGCCCGCGGAGAGCGTGTGG 360
DB	301	TTCTCTTTTCGACGCGAGGACGTCTGTAGGCGACCCGACCTGTGCCCGCGAGAGCGTGTGG 360

RESULT 6
AAX77719
ID AAX77719 standard; DNA; 549 BP.
XX
AC AAX77719;

QY 361 CACGGCTCCGACCCACGGGCGCGCCCTGACCGACAGCTACTGCGACAGCTGGCGGACG 420
 Db |||||
 QY 361 CATGGCTCGGACCCACGAGCGCGAGCTGACCGAGAGCTACTGTGAGAGCTGGCGGACG 420
 Db |||||
 QY 421 GAGGCCCGGCGGCCACCGGGGAGGGTGTGCTGCTGCTGGCGGCGAGGCTCTGAGCAG 480
 Db |||||
 QY 421 GAGGCTCCCTCGGCGACGGGCCAGGGCTCTGCTGCTGCTGGGGGCGAGGCTCTCTGGGCGAG 480
 Db |||||
 QY 481 GAGGCCCGGAGCTGCCGCCACGGCTTCGTGGTGTCTGTCATCGAGAACAGGCTCATGACC 540
 Db |||||
 QY 481 AGTCCGCGAGCTGCCATCAGCGCTACATCGTGTCTGATGAGAACAGCTTCATGACT 540
 QY 541 TCCTTC 546
 Db |||||
 QY 541 GCCTCC 546

RESULT 7
 AAZ51291
 ID AAZ51291 standard; cDNA; 549 BP.
 AC
 AAZ51291;

06-JUN-2000 (first entry)

Human angiogenesis inhibitor, endostatin cDNA.

Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;
 angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
 antiproliferative; antidiabetic; ophthalmological; immunosuppressant;
 vasotropic; vulnery; treatment; antiarteriosclerosis; tumour;
 metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
 ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 wound granulation; keloid scar; gene therapy; ss.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 1..549

FT /*tag= a

FT /product= "Endostatin"

FT /note= "Does not include stop codon"

FT /partial

FN MO200011033-A2.

PD 02-MAR-2000.

PF 25-AUG-1999; 99MO-US019329.

PR 25-AUG-1998; 98US-0097883P.

PA (LEXI-) LEXINGEN PHARM CORP.

PI Lo K, Li Y, Gillies SD;

DR WPI; 2000-237616/20.

DR P-PSDB; AAY70252.

Novel fusion protein of angiostatin or endostatin and an immunoglobulin
 FC region, useful for treating conditions mediated by angiogenesis, such
 as rheumatoid arthritis, tumors and macular degeneration.

Claim 12; Page 40-41; 68pp; English.

The patent discloses a DNA molecule encoding a fusion protein comprising
 a signal sequence, an immunoglobulin Fc region, and an angiogenesis
 inhibitor selected from angiostatin, endostatin, a plasminogen fragment
 having angiostatin activity, a collagen XVIII fragment having endostatin
 activity, or combinations of them. The fusion protein (immunofusin) is
 used to inhibit angiogenesis and to treat diseases or conditions mediated

by angiogenesis. Conditions that may be treated include solid tumours,
 blood born tumours, tumour metastasis, benign tumours including
 haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
 granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 e.g. diabetic retinopathy, retinopathy of prematurity, macular
 degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
 plaque neovascularisation, telangiectasia, haemophilic joints,
 angiofibroma, wound granulation, and excessive or abnormal stimulation of
 endothelial cells, intestinal cells, atherosclerosis, scleroderma and
 hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
 gene therapy. The present sequence is a cDNA encoding human endostatin
 used in the construction of immunofusin containing human immunoglobulin
 gamma (IgG) Fc fragment. Note: This sequence is stated in claim 12 as
 being amino acid sequence of plasminogen fragment, however, the rest of
 the specification refers to this sequence as being nucleotide sequence of
 human endostatin

XX Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match 75.3%; Score 418; DB 3; Length 549;

Best Local Similarity 85.3%; Pred. No. 8.9e-61;

Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACAGGACTTCCAGCTGGTGTGCTGCACCTGGTGGCCCTGAACAGCCCGAGCGG 60

Db 1 CACAGCCACCGCGACTTCCAGCGGTGTCTCCACTGGTGTGCTCAACAGCCCTGTCA 60

QY 61 GCGGCGATGCGAGGCAATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCGCG 120

Db 61 GCGGCGATGCGGGGCAATCCGGGCGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCGCG 120

QY 121 GGGTGTGCGCGCACTTCCGGGCTTCTGTGTGTGCGGCTGCGAGGACTCTTACAGCATC 180

Db 121 GGGTGTGCGGCGCACTTCCGGGCTTCTGTGTGTGCGGCTGCGAGGACTCTTACAGCATC 180

QY 181 GTGCGCGCGCGCGACCGCGGCGGCTGCTCACTCAGGCGAGAGTGTCTCTTC 240

Db 181 GTGCGCGCGCGCGACCGCGGCGGCTGCTCACTCAGGCGAGAGTGTCTCTTC 240

QY 241 CCCAGCTGGAGGCTTATTTCGGGCTCCGAGGCGCAGCTGAAGCCCGGGGCGCGCATC 300

Db 241 CCCAGCTGGAGGCTTATTTCGGGCTCCGAGGCTCCGAGGCGCAGCTGAAGCCCGGGGCGCGCATC 300

QY 301 TTCTCTTTTACGCGGAGAGATGCTGCGAGCACCCCGCTGCGCCCGGAGAGGCTGTGG 360

Db 301 TTCTCTTTTACGCGGAGAGATGCTGCGAGCACCCCGCTGCGCCCGGAGAGGCTGTGG 360

QY 361 CAGGCTTCCGACCCCGAGCGGCGCGCGCTGACGACAGCTACTGCGAGACGTGGCGGACG 420

Db 361 CAGGCTTCCGACCCCGAGCGGCGCGCGCTGACGACAGCTACTGCGAGACGTGGCGGACG 420

QY 421 GAGGCCCGGCGCGCACCGGGCAGGGGTGCTGCTGCTGGCGGCGAGCTGTGGAGCAG 480

Db 421 GAGGCCCGGCGCGCACCGGGCAGGGGTGCTGCTGCTGGCGGCGAGCTGTGGAGCAG 480

QY 481 GAGGCCCGGAGCTGGCGCGCACCGGCTTCTGTGGTGTCTGTGATCGAGAACAGCGTCTATGACC 540

Db 481 AGTGGCGCGAGCTGGCATCAGCGCTACATCGTGTCTGTGATTTGAGACAGCTTATGACT 540

QY 541 TCCTTC 546

Db 541 GCCTCC 546

RESULT 8

AAZ29884

ID AAA29884 standard; cDNA; 549 BP.

AC AAA29884;

DT 22-AUG-2000 (first entry)

XX

DE Human angiogenesis inhibiting factor 1 encoding cDNA.
XX
KW Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;
KW abnormal vessel disease; ss.
XX
OS Homo sapiens.
PN CN1244536-A.
XX
PD 16-FEB-2000.
XX
PF 10-AUG-1998; 98CN-00117150.
XX
PR 10-AUG-1998; 98CN-00117150.
XX
PA (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.
XX
PI Yang Z, Guo W;
XX
DR WPI; 2000-388168/34.
DR P-PSDB; AAY90771.
XX
PT Angiogenesis inhibiting factor 1 and its derivative useful for treating
PT tumors.
XX
PS Example 1; Fig 5; 41pp; Chinese.
XX
CC The present sequence encodes an angiogenesis inhibiting factor (I),
CC designated IAF-1. The present invention also describes: (1) preparation
CC of (I) and its derivative; (2) an IAF binding acceptor and its
CC preparation; and (3) an IAF antibody. (I) is useful for preparing new
CC biological preparations for effectively treating various tumours and
CC abnormal-vessel diseases. The IAF antibody is preferably a polyclonal
CC antibody, mosaic antibody, single stranded antibody and human originated
CC antibody
XX
SQ Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;
Query Match 75.3%; Score 418; DB 3; Length 549;
Best Local Similarity 85.3%; Pred. No. 8.9e-61;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACAGGACTCCAGCTGGTGTCTGACCTGGTGGCCCTGAACAGCCGCGAGCG 60
DB 1 CACAGCCACCGGACTTCCAGCGGTGCTCCACCTGGTGGCTCAGAGCCCTGTCA 60
QY 61 GCGGCGATGCGAGGATCCCGGGAGCGGACTTCCAGTGTCTCAGAGCGCGCGCGG 120
DB 61 GCGGCGATGCGGGGATCCCGGGGCGGACTTCCAGTGTCTCAGAGCGCGCGCGG 120
QY 121 GGGTGGCGGCGACCTTCCGGGCTTCTCTGCTGGGCTGCGAGCTCTACAGCATC 180
DB 121 GGGTGGCGGCGACCTTCCGGGCTTCTCTGCTGGGCTGCGAGCTCTACAGCATC 180
QY 181 GTGCGCGCGCGACCGCACCGGGGTGCGGTGCTCAACCTCAGGACGAGTGTCTTC 240
DB 181 GTGCGCGGTGCGGACCGCGGCGAGCGGTGCGGTGCTCAACCTCAGGACGAGTGT 240
QY 241 CCCAGCTGGGAGGCTTATCTCGGGCTCCGAGGCGAGCTGAAGCCCGGGCGCGCATC 300
DB 241 CCCAGCTGGGAGGCTTATCTCGGGCTCCGAGGCGAGCTGAAGCCCGGGCGCGCATC 300
QY 301 TTCTCTTTTCGACGCGAGAGATGCTCTCAGCACCCCGCTGGCCCGGAGAGCGTGTGG 360
DB 301 TTCTCTTTTCGACGCGAGAGATGCTCTCAGCACCCCGCTGGCCCGGAGAGCGTGTGG 360
QY 361 CACGCTCCGACCCCGAGCGGCGCGCTGACCGACAGCTACTGCGAGAGCTGGCGGACG 420
DB 361 CATGCTCGGACCCCGACCGGCGGAGGTGACCGAGAGCTACTGTGAGACGTGGCGGACG 420
QY 421 GAGGCGCGGCGGCGACCGGCGAGCGTGTGCTGTGGCGGCGAGCTGTGGAGCAG 480
DB 421 GAGGCTCCCTCGCGCACGGCGCAGGCTTCTCGCTGTGGGGGCGAGGCTCCTGGGCGAG 480

QY 481 GAGGCGCGGAGCTGCCGCGACGCCCTTCTGTGCTCTGATCGAGAACAGCGTCTATGACC 540
DB 481 AGTGCGCGGAGCTGCCATCACGCTTACATCGTGTCTGTGATGAGAACAGCTTATGACT 540
QY 541 TCGTTC 546
DB 541 GCGTCC 546
RESULT 9
AAS00867
ID AAS00867 standard; DNA; 549 BP.
XX
AC AAS00867;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human gene fragment encoding Endostatin(TM) protein.
XX
DE Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophilic joint; angiofibroma; wound granulation; ds.
XX
OS Homo sapiens.
XX
PH Key
FT CDS
FT Location/Qualifiers
FT 1..549
FT /tag= d
FT /product= "Endostatin(TM)"
FT /partial
FT /note= "None of the above CDSs have start or stop codons"
FT CDS
FT 1..546
FT /tag= c
FT /product= "Endostatin(TM) C-terminus minus 1"
FT /partial
FT /note= "Variant produced during fermentation reaction of Pichia pastoris harbouring an expression plasmid containing the present sequence"
FT CDS
FT 1..543
FT /tag= b
FT /product= "Endostatin(TM) C-terminus minus 2"
FT /partial
FT /note= "Variant produced during fermentation reaction of Pichia pastoris harbouring an expression plasmid containing the present sequence"
FT CDS
FT 1..540
FT /tag= a
FT /product= "Endostatin(TM) C-terminus minus 3"
FT /partial
FT /note= "Variant produced during fermentation reaction of Pichia pastoris harbouring an expression plasmid containing the present sequence"
XX
PN W0200119989-A2.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-US025166.
XX
PR 14-SEP-1999; 99US-0153698P.
XX
PA (ENTR-) ENTREMED INC.
XX
PI Liang H, Sim XL, Chang-Murad A, Zhou X, Madesen J, Boerner RJ;
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
XX

DR WPI; 2001-244802/25.
DR P-PSDB; AAU00896, AAU00897, AAU00898, AAU00899.
XX Producing Endostatin protein for treating angiogenesis mediated diseases
XX such as solid tumors, comprises recombinantly producing the protein using
PT an expression system, and recovering and purifying the protein.
PT
XX
XX Claim 6; Page 29; 67pp; English.
PS
XX The sequence encodes Human Endostatin(TM). The new method of the
CC invention is useful for producing, recovering and purifying Endostatin
CC (TM) from biological sources, such as biological fluids, tissues, cells,
CC culture media, and fermentation media. Endostatin(TM) is useful for
CC treating angiogenesis mediated diseases such as solid tumors, blood
CC borne tumors, leukemias, tumor metastases, benign tumors, e.g.
CC haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
CC e.g., diabetic retinopathy, retinopathy of prematurity, macular
CC degeneration, corneal graft rejection, neovascular glaucoma, colon
CC cancer, retrolental fibroplasia, rubeosis, Osler-Weber Syndrome,
CC myocardial angiogenesis, plaque neovascularization, telangiectasia,
CC haemophilic joints, angiofibroma, and wound granulation. Endostatin(TM)
CC is also useful for treating disease of excessive or abnormal stimulation
CC of endothelial cells such as intestinal adhesions, atherosclerosis, and
CC scleroderma and hypertrophic scars. Higher yields of more purified, and
CC biologically active Endostatin(TM) are obtained by the new method.
CC Endostatin(TM) can be stored in buffers for extended periods of time, and
CC also subjected to lyophilisation, while preserving biological activity.
CC Centrifugation of broth from fermentation steps in production is avoided,
CC preventing unwanted potential cellular lysis and contamination with
CC additional proteins, pigments, enzymes and other cellular chemicals and
CC debris
XX
SQ Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match 75.3%; Score 418; DB 5; Length 549;
Best Local Similarity 85.3%; Pred. No. 8.9e-61;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 1 CACACCCACAGGACTTCAGCTGTGTGTCACCTGTGTGCTGTGACACAGCCCGCAGCG 60
DB 1 CACAGCCACCGACTTCACGCGGTGTCTCCACCTGTGTGTGCTGTGCTGTGCTGTCA 60
QY 61 GCGCGCTGCGAGGATCCGGGAGCGGACTTCAGTGTCTTCAGAGGCGCGCGCGG 120
DB 61 GCGCGATGCGGGGATCCGGGCGCGACTTCAGTGTGTGCTGTGCTGTGCTGTGCTG 120
QY 121 GCGCTGGCGGCACTTCGCGGCTTCTGTGTGTGCTGTGCTGTGCTGTGCTGTGCTG 180
DB 121 GCGCTGGCGGCACTTCGCGGCTTCTGTGTGTGCTGTGCTGTGCTGTGCTGTGCTG 180
QY 181 GTGCGCGCGCGCGACCGGAGCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 240
DB 181 GTGCGCGCGCGCGACCGGCGCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 240
QY 241 CCGAGTGGGAGGCTTATTCGCGGCTTCGAGGCGGAGCTGAAGCCCGGCGCGGCGATC 300
DB 241 CCGAGTGGGAGGCTTATTCGCGGCTTCGAGGCTTCGAGGCTGAAGCCCGGCGCGG 300
QY 301 TTCTCTTTTCAGCGGAGAGATGCTCTGCGACACCCCGCTGCGCCCGGAGAGCGTGTGG 360
DB 301 TTCTCTTTTCAGCGGAGAGATGCTCTGCGACACCCCGCTGCGCCCGGAGAGAGCGTGTGG 360
QY 361 CAGCGTCCGACCCAGCGGCGCGCTGTGACCGAGCTACTGCGAGCGTGTGCGGACG 420
DB 361 CATGGCTCGGACCCCAACCGGCGCGAGGCTGTGACCGAGAGCTACTGTGAGACGTGCGGACG 420
QY 421 GAGGCGCGCGCGCGACCGCGGAGGCGTGTGCTGTGCTGTGCGCGGCGAGCTGTGAGACG 480
DB 421 GAGGCTCCCTCGGCGACCGGCGCGAGGCTGTGCTGTGCTGTGCGGCGAGCTGTGAGACG 480
QY 481 GAGGCGCGAGTGTGCGCGCGCGCTTCTGTGTGTGCTGTGCTGTGCTGTGCTGTGCTG 540
DB 481 GAGGCGCGAGTGTGCGCGCGCGCTTCTGTGTGTGCTGTGCTGTGCTGTGCTGTGCTG 540

Db 481 AGTGCGCGAGCTGCCATCAGCGCTACATCGTGTCTGTGATTGAGACAGCTTCATGACT 540
QY 541 TCCTTC 546
DB 541 GCCTCC 546
RESULT 10
ABA00774
ID ABA00774 standard; cDNA; 549 BP.
XX
AC ABA00774;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human endostatin coding sequence.
XX
KW Gene; human; plasminogen; angiotensin; angiotensin; neovascularisation;
KW kringle domain; cell proliferation; viral vector; replication-defective;
KW cancer; tumour; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..549
FT /tag= a
FT /product= "Endostatin"
XX
PN WO200288173-A2.
XX
PD 07-NOV-2002.
XX
PF 29-APR-2002; 2002WO-US013461.
XX
PR 30-APR-2001; 2001US-0287673P.
PR 05-APR-2002; 2002US-0370634P.
XX
PA (CELL-) CELL GENESYS INC.
XX
PI Chang B, Wu WW, Macarthur J, Patel S, Jooss K, Mendez M;
XX
DR WPI; 2003-129131/12.
DR P-PSDB; AAG79753.
XX
PT New recombinant viral vector expressing human angiotensin useful for
PT inhibiting angiogenesis in a mammalian subject with cancer or tumor.
XX
PS Example 4; Page 80-82; 83pp; English.
XX
CC This sequence encodes endostatin. Endostatin is a 20 kD C-terminal
CC fragment of collagen XVII that inhibits angiogenesis. The endostatin
CC coding sequence may be used in the recombinant viral vector of the
CC invention for obtaining angiotensin activity. The vector comprises a
CC promoter capable of expressing human angiotensin operably linked to a
CC structural gene encoding one or more domains of human angiotensin. The
CC vector, which may be a replication-defective viral vector, is useful for
CC inhibiting angiogenesis in a mammal, especially with cancer or a tumor.
CC Note: This sequence is given incorrectly in the sequence listing of the
CC specification as an amino acid sequence
XX
SQ Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match 75.3%; Score 418; DB 7; Length 549;
Best Local Similarity 85.3%; Pred. No. 8.9e-61;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 1 CACACCCACAGGACTTCAGCTGTGTGTCACCTGTGTGCTGTGACACAGCCCGCAGCG 60
DB 1 CACAGCCACCGACTTCACGCGGTGTCTCCACCTGTGTGCTGTGCTGTGCTGTGCTG 60
QY 61 GCGCGCATCCGAGGCTATCCGGGAGCGGACTTCAGTGTCTTCAGTGTCTTCAGAGGCGCGCGCG 120
DB 61 GCGCGCATCCGAGGCTATCCGGGAGCGGACTTCAGTGTCTTCAGTGTCTTCAGAGGCGCGCGCG 120

CC combinations of angiostatin, endostatin, interferon, thrombospondin,
CC interferon-inducible protein and platelet factor 4, and have anti-
CC angiogenic and/or anti-tumor activity. The multifunctional protein may
CC exhibit useful properties such as having similar or greater biological
CC activity when compared to a single factor or by having improved half-life
CC or decreased adverse side effects, or a combination of these properties.
CC The proteins can be used for treating an angiogenic-mediated disease,
CC e.g. cancer, diabetic retinopathy, macular degeneration, or arthritis.
CC They can also be used for inhibiting the production of tumor cells
CC (characteristic of lung, breast, ovarian, prostate, pancreatic, gastric,
CC colon, renal, bladder cancers; melanoma, hepatoma, sarcoma and lymphoma)
CC in a patient and for inhibiting tumor growth. The present sequence is
CC used in the course of the invention
CC
SQ Sequence 552 BP; 83 A; 196 C; 179 G; 94 T; 0 U; 0 Other;

Query Match 75.3%; Score 418; DB 2; Length 552;
Best Local Similarity 85.3%; Pred. NO. 8.8e-61;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACCCAGGACTTCAGCTGTGTGTCACCTGTGGCCCTGAACAGCCCGCAGCCG 60
DB 1 CACAGCCACCGGACTTCAGCCGGTGTCTCCACCTGTGGTTCACAGCCCGCTGTCA 60
QY 61 GCGCGATGCGAGGATCCGGGAGCGGACTTCAGTGTCTTCCAGCAGCGCGCGCGCG 120
DB 61 GCGCGATGCGAGGATCCGGGAGCGGACTTCAGTGTCTTCCAGCAGCGCGCGCGCG 120
QY 121 GAGCTGGCGCGGACCTTCCTGGGCTTCCTGCTGCGGCTGAGGACCTCTACAGCATC 180
DB 121 GAGCTGGCGCGGACCTTCCTGGGCTTCCTGCTGCGGCTGAGGACCTCTACAGCATC 180
QY 181 GTGCGCGCGCGGACCTTCCTGGGCTTCCTGCTGCGGCTGAGGACCTCTACAGCATC 240
DB 181 GTGCGCGCGCGGACCTTCCTGGGCTTCCTGCTGCGGCTGAGGACCTCTACAGCATC 240
QY 241 CCCAGCTGGAGGCTTATTCCTGGGCTTCCTGAGGCTGAGGCTGAGGCTGAGGCT 300
DB 241 CCCAGCTGGAGGCTTATTCCTGGGCTTCCTGAGGCTGAGGCTGAGGCTGAGGCT 300
QY 301 TTCTCTTTTCAGCGCAGAGATGCTCTGAGCAGACCCCGCTGCGGCGGAGAGAGCTGTGG 360
DB 301 TTCTCTTTTCAGCGCAGAGATGCTCTGAGCAGACCCCGCTGCGGCGGAGAGAGCTGTGG 360
QY 361 CAGGCTCCGACCCGAGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 420
DB 361 CAGGCTCCGACCCGAGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 420
QY 421 GAGGCGCGCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 480
DB 421 GAGGCTCCCTCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 480
QY 481 GAGGCGCGGAGCTGCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 540
DB 481 AGTGCGCGGAGCTGCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTTCACT 540
QY 541 TCCTTC 546
DB 541 GCCTCC 546

RESULT 12
AAAG8203
ID AAA68203 standard; cDNA; 552 BP.
XX
XX AAA68203;
AC
XX
XX 27-OCT-2000 (first entry)
DT
XX
XX Human endostatin cDNA sequence.
DE
XX
XX Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;
KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
CC

QY 121 GGGCTGGCGGACCTTCGGGCTTCCTGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 180
DB 121 GGGCTGGCGGACCTTCGGGCTTCCTGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 180
QY 181 GTGCGCGCGGCGGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 240
DB 181 GTGCGCGCGGCGGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 240
QY 241 CCAGCTGGAGGCTTATTCCTGGGCTTCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 300
DB 241 CCAGCTGGAGGCTTATTCCTGGGCTTCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 300
QY 301 TTCTCTTTTCAGCGCAGAGATGCTCTGAGCAGACCCCGCTGCGGCGGAGAGAGCTGTGG 360
DB 301 TTCTCTTTTCAGCGCAGAGATGCTCTGAGCAGACCCCGCTGCGGCGGAGAGAGCTGTGG 360
QY 361 CAGGCTCCGACCCGAGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 420
DB 361 CATGGCTCGGACCCGAGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 420
QY 421 GAGGCGCGGCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 480
DB 421 GAGGCTCCCTCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTTCACT 480
QY 481 GAGGCGCGGAGCTGCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 540
DB 481 AGTGCGCGGAGCTGCGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTTCACTGACT 540
QY 541 TCCTTC 546
DB 541 GCCTCC 546

RESULT 12
AAAG8203
ID AAA68203 standard; cDNA; 552 BP.
XX
XX AAA68203;
AC
XX
XX 27-OCT-2000 (first entry)
DT
XX
XX Human endostatin cDNA sequence.
DE
XX
XX Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;
KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
CC

psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;
cerebral collateral; arteriovenous malformation; rubecosis; cancer;
diabetic retinopathy; arthritis; wound healing; peptic ulcer;
Helicobacter related disease; fracture; cat scratch fever; ss.
Homo sapiens.
WO2000032631-A2.
08-JUN-2000.
06-DEC-1999; 99WO-US028897.
04-DEC-1998; 98US-00206059.
(ENTR-) ENTREMED INC.
Macdonald NJ, Sim KL;
WPI; 2000-412290/35.
New angiogenesis-inhibiting protein receptors, useful in methods for
treating diseases and processes that are mediated by angiogenesis, such
as solid tumors, psoriasis, scleroderma and myocardial angiogenesis.
Disclosure; Page 86; 100pp; English.
This invention relates to angiogenesis-inhibiting protein receptors, and
the DNA sequences encoding them. Angiogenesis is the generation of new
blood vessels into a tissue, and normally occurs in wound healing, foetal
and embryonal development and the formation of the corpus luteum.
CC endometrium and placenta. Angiostatin is a protein (see AAB16450 and
AAAG8202) involved in angiogenesis, and has an amino acid sequence
similar to that of a plasminogen fragment (see murine plasminogen
AAB16450). Angiostatin has the ability to inhibit angiogenesis.
CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and
AAAG8203). Sequences AAG68242 and AAB16522 represent coding and protein
sequences of human laminin. Laminin is an angiostatin binding protein,
and some of the peptides of the invention share homology with regions of
laminin. Peptides AAB16452-B16521 (excluding AAB16450) are the
angiogenesis-inhibiting protein receptor fragments of the invention. The
peptides bind either angiostatin or endostatin and can be used in methods
for treating diseases and processes that are mediated by angiogenesis,
such as solid tumors, psoriasis, scleroderma, myocardiocard angiogenesis,
Crohn's disease, cerebral collateral, arteriovenous malformations,
rubecosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
Helicobacter related diseases, fractures, placenta and cat scratch
fever. They are useful for the detection and prognosis of cancer. DNA
sequences A628204-A628241 encode the peptides of the invention
Sequence 552 BP; 83 A; 196 C; 179 G; 94 T; 0 U; 0 Other;
Query Match 75.3%; Score 418; DB 3; Length 552;
Best Local Similarity 85.3%; Pred. No. 8.8e-61;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 1 CACACCCACGAGGACTTCAGCTGTGTGTCACCTGTGGCCCTGAACAGCCCGACGCG 60
DB 1 CACAGCCACCGGACTTCAGCCGGTGTCTCAGCTGTGTGGCTCAACAGCCCTGTCA 60
QY 61 GCGCGATCGAGGATCGGCGGAGCGGACTTCAGTGTCTTCAGGAGCGCGCGCG 120
DB 61 GCGCGATCGGCGGATCGGCGGCGGCGGACTTCAGTGTCTTCAGGAGCGCGCGCG 120
QY 121 GCGTGGCGGCGGACTTCGCGGCGCTTCGTCGCGGCTGAGGACCTCTACAGCATC 180
DB 121 GCGTGGCGGCGGACTTCGCGGCGCTTCGTCGCGGCTGAGGACCTCTACAGCATC 180
QY 181 GTGCGCGCGCGGACCGACCGACCGGCGTGCCTGTGTCACCTTCAGGAGGCTCTTC 240
DB 181 GTGCGCGCGGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
QY 241 CCGAGCTGGAGGCGCTTATTCGCGGCTCCGAGGCGGCGGCGGCGGCGGCGGCG 300

Db 241 CCCAGCTGGGAGGCTCTGTTCAGGCTCTGAGGCTCGCTGAGCCCGGGCAGGCATC 300
QY 301 TTCTCTTTTCGACGCGCAGAGATGTCTTGACGACACCCCGCTGCCCCGGAAGAGCGTGTGG 360
Db 301 TTCTCTTTTCGACGCGCAGAGATGTCTTGAGGACACCCCTGAGGCGGCGGCGGCGTGTGG 360
QY 361 CACGGTCTCCGACCCCGAGCGGGCGCCCTGACGACAGCTACTGCGAGCGTGGCGGACG 420
Db 361 CATGGCTCGGACCCCAACCGGCGGAGGCTGACCGAGAGCTACTGTGAGACGTGGCGGACG 420
QY 421 GAGGCGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 421 GAGGCTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 481 GAGGCGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 481 AGTGCGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
QY 541 TCCTTC 546
Db 541 GCCTCC 546
RESULT 13
AAC62023
ID AAC62023 standard; DNA; 552 BP.
XX
AC AAC62023;
XX
DT 06-MAR-2001 (first entry)
XX
DE Nucleotide sequence of human endostatin encoded by plasmid pMALCH#15.
XX
KW Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
KW endostatin; cancer; tumour growth; angiogenesis; ss.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..552
FT /tag= a
FT /product= "endostatin"
XX
FN WO2000060945-A1.
XX
PD 19-OCT-2000.
XX
PF 12-APR-2000; 2000WO-US009747.
XX
PR 13-APR-1999; 99US-0129084P.
XX
PA (MERI) MERCK & CO INC.
XX
PI Desanti CL, Strchl WR;
XX
DR WPI; 2000-686970/67.
DR P-PSDB; AAB30493.
XX
PT Preparation of soluble recombinant endostatin involves transforming
PT Streptomycete host with expression vector comprising nucleotide sequence
PT encoding endostatin operably linked to linker and leader peptide.
XX
PS Example 1; Fig 6; 57pp; English.
XX
CC The present sequence encodes human endostatin. The protein is expressed
CC in Streptomyces. Leader sequences of Streptomyces sp. strain C5 SnpA and
CC S. venezuelae alpha-amylase proteins are linked to the N-terminal of
CC endostatin. This ensures that endostatin protein is produced as a
CC secreted, soluble protein which needs no refolding, is stable in the
CC fermentation broth and is produced in large quantities. The method is
CC used for preparing soluble recombinant human, murine or primate
CC endostatin, which is useful in the treatment of cancer, inhibition of

Age Group	Percentage
18-24	10
25-34	35
35-44	25
45-54	15
55-64	10
65-74	5
75-84	2
85+	1

1	418	75.3	546	4	US-09-315-689-4	Sequence 4, Appl
2	418	75.3	552	3	US-09-206-059-30	Sequence 30, Appl
3	418	75.3	553	4	US-09-231-077D-9	Sequence 9, Appl
4	418	75.3	3394	1	US-08-159-784-4	Sequence 4, Appl
5	410.8	74.0	534	4	US-09-315-689-6	Sequence 6, Appl
6	375.8	67.7	558	3	US-09-449-293-3	Sequence 3, Appl
7	375.8	67.7	558	4	US-09-775-325-3	Sequence 3, Appl
8	375.8	67.7	558	4	US-09-775-174-3	Sequence 3, Appl
9	375.8	67.7	565	3	US-08-985-526-37	Sequence 37, Appl
10	374.2	67.4	4031	1	US-08-159-784-1	Sequence 1, Appl
11	373.4	67.3	573	4	US-09-561-500-12	Sequence 12, Appl
12	373.4	67.3	573	4	US-09-561-108-12	Sequence 12, Appl
13	373.4	67.3	573	4	US-09-561-526-12	Sequence 12, Appl
14	373.4	67.3	573	4	US-09-561-499-12	Sequence 12, Appl
15	373.4	67.3	573	4	US-09-998-831-12	Sequence 12, Appl
16	373.2	67.2	568	4	US-09-231-077D-8	Sequence 8, Appl
17	373.2	67.2	580	4	US-09-231-077D-7	Sequence 7, Appl
18	371.6	67.0	582	4	US-09-231-077D-6	Sequence 6, Appl
19	370.8	66.8	620	4	US-09-231-077D-5	Sequence 5, Appl
20	60.8	11.0	1116	4	US-09-252-991A-2102	Sequence 2102, Ap
21	60.8	11.0	1294	4	US-09-252-991A-1793	Sequence 1793, Ap
22	59.4	10.7	1200	4	US-09-252-991A-1861	Sequence 1861, Ap
23	59.4	10.7	1722	4	US-09-252-991A-2173	Sequence 2173, Ap
24	57.4	10.3	741	4	US-09-252-991A-6774	Sequence 6774, Ap
25	57.4	10.3	1557	4	US-09-252-991A-6920	Sequence 6920, Ap
26	57.4	10.3	1590	4	US-09-252-991A-6824	Sequence 6824, Ap
27	55.8	10.1	1820	1	US-08-173-508-7	Sequence 7, Appl

1	CACACCCACGAGACTTCAGCTGGTGTGCTACCTGGTGGCCCTGTAACAGCCCCCGACGCG	60	Qy	
	1	CACAGCCACCGGACTTCAGCGGGTCTCCACCTGGTTGGGCTCAACAGAGCCCCCTGTCA	60	Db
61	GGCGGATGCGAGGATCCGGGGACGGGACTTCAGTGTCTCCAGAGCGGCGCGCGCG	120	Qy	
	61	GGCGGCATGGGGGCATCCGCGGGCCGACTTCAGTGTCTCCAGAGCGCGGGCCGTG	120	Db
121	GGGTGTGCGGCACTTCGCGGGCCCTTCCTGTGCTCGCGGCTGCAGGACTCTACAGCATC	180	Qy	
	121	GGGTGTGCGGGCACTTCGCGGCCCTTCCTGTGCTCGCGCTGCAGGACTCTACAGCATC	180	Db
181	GTGCGCGCGCGACCGCACCGGGGTGCCGTGCTCAACTCAGGAGCAGGGTGCCTTTC	240	Qy	
	181	GTGCGCGGTCCGACCGCGCAGCGGTGCCCACTGCTCAACTCAGGAGCAGACTCTGCTTT	240	Db
241	CCAGCTGGAGGCGCTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC	300	Qy	
	241	CCAGCTGGGAGGCTCTGTTCAGGCTCTGAGGGTCCGCTGAAGCCCCGGGGCAGCATC	300	Db
301	TTCTCTTTGACGGCAGAGATGCTCTGACGACCCCGGCTTGCCCCCGGAAGAGGCTGTGG	360	Qy	
	301	TTCTCTTTGACGGCAGAGCTCTCTGAGGCACCCCACTTGCCCCCAGAAGAGCGTGTGG	360	Db
361	CACGGCTCCGACCCGACGGGGCGCCCTGTGACCGACAGCTACTGTGGAGACGTGGCGGACG	420	Qy	
	361	CATGGCTCGGACCCCAACGGCGCAGGCTGACCGAGAGCTACTGTGAGACGTGGCGGACG	420	Db

RESULT 6

US-09-449-293-3
; Sequence 3, Application US/09449293
; Patent No. 6267954
; GENERAL INFORMATION:
; APPLICANT: Aditbol, Marc
; APPLICANT: Uteza, Yves
; APPLICANT: Menasche, Maurice
; APPLICANT: Bossard, Carine
; APPLICANT: Van Den Bergh, Loic
; APPLICANT: Bonnel, Sebastien
; APPLICANT: Prats, Herve
; APPLICANT: Honiger, Jiri
; APPLICANT: Neuner-Jehle, Martin
; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
; FILE REFERENCE: 8076.202US01
; CURRENT APPLICATION NUMBER: US/09/449,293
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Rattus rattus
US-09-449-293-3

Query Match 67.7%; Score 375.8; DB 3; Length 558;
Best Local Similarity 79.8%; Pred. No. 3e-66;
Matches 443; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY	1	CACACCACAGGACTTCAGCTGGTGGCTGCACTGGTGGCCCTGAACAGCCCGCAGCG	60
DB	4	CATACATCAGGACTTTCAGCCAGTGTCCACCTGGTGGCACTGAACACCCCTGTCT	63
QY	61	GGCGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGAGCGCGCGCG	120
DB	64	GGAGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGTCTCCAGAGCGCGCG	123
QY	121	GGGTGCGCGGCACTTCCGGGCTTCTGTGCTGGCGCTGCGAGCTCTACAGCATC	180
DB	124	GGGTGCTGCGGCACTTCCGGGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG	183
QY	181	GTGCGCGCGCGCAGCAGCCGGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG	240
DB	184	GTGCGCGCTGCTGACCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	243
QY	241	CCAGCTGGAGGCTTTATTTCTGGGCTCCGAGGGCGAGCTGAAGCCCGGCGCATC	300
DB	244	CCAGCTGGGACTCCCTGTTTCTGGCTCCAGGGTCAACTGCAACCCGGGCGCGCATC	303
QY	301	TTCTCTTTCCAGCGCAGAGATGCTCTGAGCAGCCCGCTGCGCCCGGAGAGCGTGG	360
DB	304	TTTTCTTTTTCAGCGCAGAGATGCTCTGAGCAGCCCGCTGCGCCCGGAGAGCGTGG	363
QY	361	CACGGCTCCGACCCCGAGCGGCGCGCTGAGCAGCTGCTGAGAGCTGCGGAGCG	420
DB	364	CACGGCTCGGACCCCGAGCGGCGCGCTGAGCAGCTGCTGAGAGCTGCGGAGCG	423
QY	421	GAGGCCCGCGCGCAGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480
DB	424	GAACTACTGGGCTACAGTCAAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	483
QY	481	GAGGCCGAGCTGCGCCCGCAGCGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG	540
DB	484	AAAGCTCGAGCTGCGCCAGCAGCTACATCGTCTGTGCTGCTGCTGCTGCTGCTGCT	543
QY	541	TCCTTCTCCCAAGTAG	555
DB	544	TCCTTCTCCCAAGTAG	558

RESULT 7

US-09-775-325-3
; Sequence 3, Application US/09775325
; Patent No. 6500449
; GENERAL INFORMATION:
; APPLICANT: Aditbol, Marc
; APPLICANT: Uteza, Yves
; APPLICANT: Menasche, Maurice
; APPLICANT: Bossard, Carine
; APPLICANT: Van Den Bergh, Loic
; APPLICANT: Bonnel, Sebastien
; APPLICANT: Prats, Herve
; APPLICANT: Honiger, Jiri
; APPLICANT: Neuner-Jehle, Martin
; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
; FILE REFERENCE: 8076.202US01
; CURRENT APPLICATION NUMBER: US/09/775,325
; CURRENT FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: US 09/449,293
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Rattus rattus
US-09-775-325-3

Query Match 67.7%; Score 375.8; DB 4; Length 558;
Best Local Similarity 79.8%; Pred. No. 3e-66;
Matches 443; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY	1	CACACCACAGGACTTCAGCTGGTGGCTGCACTGGTGGCCCTGAACAGCCCGCAGCG	60
DB	4	CATACATCAGGACTTTCAGCCAGTGTCCACCTGGTGGCACTGAACACCCCTGTCT	63
QY	61	GGCGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGAGCGCGCGCG	120
DB	64	GGAGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGTCTCCAGAGCGCGCG	123
QY	121	GGGTGCGCGGCACTTCCGGGCTTCTGTGCTGGCGCTGCGAGCTCTACAGCATC	180
DB	124	GGGTGCTGCGGCACTTCCGGGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG	183
QY	181	GTGCGCGCGCGCAGCAGCCGGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG	240
DB	184	GTGCGCGCTGCTGACCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	243
QY	241	CCAGCTGGAGGCTTTATTTCTGGGCTCCGAGGGCGAGCTGAAGCCCGGCGCATC	300
DB	244	CCAGCTGGGACTCCCTGTTTCTGGCTCCAGGGTCAACTGCAACCCGGGCGCGCATC	303
QY	301	TTCTCTTTCCAGCGCAGAGATGCTCTGAGCAGCCCGCTGCGCCCGGAGAGCGTGG	360
DB	304	TTTTCTTTTTCAGCGCAGAGATGCTCTGAGCAGCCCGCTGCGCCCGGAGAGCGTGG	363
QY	361	CACGGCTCCGACCCCGAGCGGCGCGCTGAGCAGCTGCTGAGAGCTGCGGAGCG	420
DB	364	CACGGCTCGGACCCCGAGCGGCGCGCTGAGCAGCTGCTGAGAGCTGCGGAGCG	423
QY	421	GAGGCCCGCGCGCAGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480
DB	424	GAACTACTGGGCTACAGTCAAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	483
QY	481	GAGGCCGAGCTGCGCCCGCAGCGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG	540
DB	484	AAAGCTCGAGCTGCGCCAGCAGCTACATCGTCTGTGCTGCTGCTGCTGCTGCTGCT	543
QY	541	TCCTTCTCCCAAGTAG	555
DB	544	TCCTTCTCCCAAGTAG	558

RESULT 8

US-09-775-174-3
; Sequence 3, Application US/09775174
; Patent No. 6663894
; GENERAL INFORMATION:
; APPLICANT: Abitbol, Marc
; APPLICANT: Uteza, Yves
; APPLICANT: Menasche, Maurice
; APPLICANT: Roseard, Carine, Loic
; APPLICANT: Van Den Bergh, Loic
; APPLICANT: Bonnel, Sebastian
; APPLICANT: Frats, Hervé
; APPLICANT: Honiger, Jiri
; APPLICANT: Neuner-Jehle, Martin
; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
; FILE REFERENCE: 8075 202USD2
; CURRENT APPLICATION NUMBER: US/09/775,174
; CURRENT FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: US 09/449,293
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Rattus rattus
US-09-775-174-3

Query Match 67.7%; Score 375.8; DB 4; Length 558;
Best Local Similarity 79.8%; Pred. No. 3e-66;
Matches 443; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY	1	CACACCCACGAGCTTCAGTGTGTGTCAGCTGTGCTGCGCTGCTGACAGCCGCGAGCGG 60
Db	4	CATCTCATCAGGACTTTCCAGCAAGTGTCTCCACCTGTGCTGACAGCCGCGAGCGG 63
QY	61	GGCGGATCGGAGGATCCGGGAGCGGACTTCAGTGTGCTGCGGAGGCGCGCGCGG 120
Db	64	GGAGGATCGGAGGATCCGGTGGAGGATTTCCAGTGTGCTGCGGAGGCGCGCGG 123
QY	121	GGGCTGGCGGACCTTCCTGGGCTTCCTGCTGCTGCGGCTGCGGAGGCTTCACATC 180
Db	124	GGGCTGGCGGACCTTCCTGGGCTTCCTGCTGCTGCGGCTGCGGAGGCTTCATATC 183
QY	181	GTGCGCGCGGCGACCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db	184	GTGCGCGCGGCGACCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
QY	241	CCGAGCTGGAGGCTTATTCCTGGGCTTCAGGCGGAGGCTGAGCCCGGCGCGGATC 300
Db	244	CCGAGCTGGAGGCTTATTCCTGGGCTTCAGGCGGAGGCTGAGCCCGGCGCGGATC 303
QY	301	TTCTCTTTTCGAGCGGAGGATGCTGCTGAGCACCGGCTGCTGCTGCTGCTGCTGCT 360
Db	304	TTTCTTTTTCGAGCGGAGGATGCTGCTGAGCACCGGCTGCTGCTGCTGCTGCTGCT 363
QY	361	CAGGCTCGGACCCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db	364	CAGGCTCGGACCCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
QY	421	GAGGCGCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db	424	GAATCTACTGGGCTTACAGGTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 483
QY	481	GAGGCGGAGGCTGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db	484	AAAGCTGGAGGCTGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
QY	541	TCCTCTCTCAAGTAG 555
Db	544	TCCTCTCTCAATAG 558

RESULT 9

US-08-985-526-37
; Sequence 37, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr. Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-985-526-37

Query Match 67.7%; Score 375.8; DB 3; Length 565;
Best Local Similarity 79.8%; Pred. No. 3e-66;
Matches 443; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY	1	CACACCCACGAGCTTCAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
Db	11	CATCTCATCAGGACTTTCCAGCAAGTGTCTCCACCTGTGCTGCTGCTGCTGCTGCT 70
QY	61	GGCGGATCGGAGGATCCGGGAGCGGACTTCAGTGTGCTGCTGCTGCTGCTGCTGCT 120
Db	71	GGAGGATCGGAGGATCCGGTGGAGGATTTCCAGTGTGCTGCTGCTGCTGCTGCTGCT 130
QY	121	GGGCTGGCGGACCTTCCTGGGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Db	131	GGGCTGTGCGGACCTTCCTGGGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 190
QY	181	GTGCGCGCGCGGACCGGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db	191	GTGCGCGCGCGGACCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 250
QY	241	CCGAGCTGGAGGCTTATTCCTGGGCTTCAGGCGGAGGCTGAGCCCGGCGCGGATC 300
Db	251	CCGAGCTGGAGGCTTATTCCTGGGCTTCAGGCGGAGGCTGAGCCCGGCGCGGATC 310
QY	301	TTCTCTTTTCGAGCGGAGGATGCTGCTGAGCACCGGCTGCTGCTGCTGCTGCTGCTGCT 360
Db	311	TTTCTTTTTCGAGCGGAGGATGCTGCTGAGCACCGGCTGCTGCTGCTGCTGCTGCTGCT 370
QY	361	CAGGCTCGGACCCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db	371	CACGCTCGGACCCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430

QY 421 GAGGCCCGCGCCACCGGCGAGCGCTCGTCTGCGGGGCGAGGCTGCTGGAGCAG 480
Db 431 GAACTACTGGGGCTACAGGTACAGCTTCTCCCTGCTGTACAGCAGGCTCTCTGGACAG 490
QY 481 GAGGCCCGAGCTGCGCCACGCTTCTGTGTCTGTCATCGAGAACAGCGTCATGACC 540
Db 491 AAAGTGTGAGCTGCGCACACAGCTACATCTGCTGTGCTATGAGATAGCTTTCATGACC 550
QY 541 TCCTTCTCCAACTAG 555
Db 551 TCTTCTCCAAATAG 565

RESULT 10

US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 555X
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4031
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-159-784-1

Query Match 67.4%; Score 374.2; DB 1; Length 4031;
Best Local Similarity 79.8%; Pred. No. 78-66; Indels 0; Gaps 0;
Matches 442; Conservative 0; Mismatches 113;
QY 1 CACACCCACAGGACTTCCAGTGTGTGTCACCTGGTGGCCCTGAAACAGCCCGCAGCG 60
Db 3313 CATACTCATCAGGACTTTCAGCCAGTGTCTCCACTGGTGGCACTGAAACACCCCTGTCT 3372
QY 61 GCGGGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGGCGCGCGCG 120
Db 3373 GGAGGCATGCGTGTATCCGTGGAGCAGATTTCCAGTGTCTCCAGCAGCCGAGCGTG 3432
QY 121 GGGCTGGCGGACCTTCCGGGCTTCTGTGTCTGCTGCGGCTGCGAGGACTTACAGATC 180
Db 3433 GGGCTGTGCGGACCTTCCGGGCTTCTGTGTCTGCTGCGGCTGCGAGGACTTCTATAGCATC 3492

QY 181 GTGCGCGCGCCGACCCGACCCGCGGTGCCCGTGTCAAACCTCAGGGACGAGGTGCTCTTC 240
Db 3493 GTGCGCGTGTGACCGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCT 3552
QY 241 CCAGCTGGGAGGCTTATTCTCGGCTCCGAGGGCCAGTGAAGCCCGGGGCCCGCATC 300
Db 3553 CCAGCTGGGACTCTCTGTGTTTCTGCTCCAGGTCAGTGAACCCCGGGGCCCGCATC 3612
QY 301 TTCTTTTCAGCGCAGAGATGCTCTGACGACCCCGCTGGCCCCCGGAGAGCGTGTGG 360
Db 3613 TTTTCTTTTGAAGGACAGATGCTCTGAGACACCCAGCCCTGGCCGAGAGCGGTATGG 3672
QY 361 CACGGCTCCGACCCCGAGCGGCGCCCTGACGACGAGCTACTGCGAGAGCTGGCGGAG 420
Db 3673 CACGGCTCGGACCCCGAGTGGGCGGAGGCTGATGAGAGTTACTGTGAGCATGGCGAAT 3732
QY 421 GAGGCCCGCGCGCCACCCGCGGCGTCTGCTGCTGTGCGCGGCGAGGCTGTGGAGCAG 480
Db 3733 GAACTACTGGGCTACAGGTCAAGGCTCTCTGCTGTGAGGCGAGGCTCTCTGGAACAG 3792
QY 481 GAGGCCGAGCTCCCGCCAGCGCTTGTGTGTCTGTGATCGAGAACAGCGTCATGACC 540
Db 3793 AAAGTGTGAGCTGCGCACACAGCTACATCTGCTGTGCTTGAAGATAGCTTCTATGACC 3852
QY 541 TCCTTCTCCAACTAG 555
Db 3853 TCTTCTCCAAATAG 3867

RESULT 11

US-09-561-500-12
; Sequence 12, Application US/09561500
; Patent No. 6342319
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)...(573)
; US-09-561-500-12

Query Match 67.3%; Score 373.4; DB 4; Length 573;
Best Local Similarity 79.9%; Pred. No. 8.9e-66;
Matches 440; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 1 CACACCCACAGGACTTCCAGTGTGTGTCACCTGGTGGCCCTGAAACAGCCCGCAGCG 60
Db 22 CATACTCATCAGGACTTTCAGCCAGTGTCTCCACTGGTGGCACTGAAACACCCCTGTCT 81
QY 61 GCGGGCATGCGAGCATCCGGGAGCGGACTTCCAGTGTCTCCAGCAGGCGCGCGCGCG 120
Db 82 GGAGGCATGCGTGTATCCGTGGAGCAGATTTCCAGTGTCTCCAGCAGCCGAGCGCGTG 141
QY 121 GGGCTGGCGGACCTTCCGGGCTTCTGTGTCTGCTGCGGCTGCGAGGACTTACAGATC 180
Db 142 GGGCTGTGCGGACCTTCCGGGCTTCTGTGTCTGCTGCGGCTGCGAGGACTTCTATAGCATC 201
QY 181 GTGCGCGCGCGCCGACCCGACCCGCGGTGCCCGTGTCAACCTCAGGGACGAGGTGCTCTTC 240

Db	262	CCGAGCTGGAGACTCCCTCTTTTCTGGCTCCAGGGTCAACTGGCAAACCCGGGGGCCCGCATC	321
Qy	301	TTCTCTTTTGGACGGCAGAGATGTCTCTGCAGCACCCCGCTGGCCCGGGAAGAGCGTGTGG	360
Db	322	TTTTCTTTTGGCGCAGAGATGTCTCTGAGACACCCAGCCTGGCCGAGGAAGCGGTATGG	381
Qy	361	CACGACTCCGACCCACAGCGGGCGCGCTGACGCAGACAGTACTGCAGAGCGTGGCGGACG	420
Db	382	CACGACTCGACCCACAGTGGCGGAGGCTGATGGAGAGTTACTTGTGAGCAATGGCGAAT	441
Qy	421	GAGGCCCGCGGGCCACCGGCAGGCGTCTGTCTGTGGCGGCAGGCTGTCTGAGCAG	480
Db	442	GAACAATACTGGGGTTACAGGTTCAGGCCCTCTCTCTGTGTACGGCAGGCTCTCTGAAACAG	501
Qy	481	GAGGCGCGAGCTGCGCGGCACGCCCTTCTGGTGTCTGTGCATCGAGAACAGCGTCATGACC	540
Db	502	AAAGTCGAGCTGCCACCAACAGGTACATCGTCTCTGTGCATTGGAATAGCTTCATGACC	561
Qy	541	TCCTTCTCCAA	551
Db	562	TCCTTCTCCAA	572

RESULT 14

```

RESULT 14
US-09-561-499-12
; Sequence 12, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDS
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-561-499-12

```

Query Match

Query Match	67.3%	Score 373.4	DB 4	Length 573
Best Local Similarity	79.9%	Prod. No. 8.se-66		
Matches 440	Conservative 0	Mismatches 111	Indels 0	Gaps 0
QY	1	CACACCCACAGAGACTCCAGACTGGTGTCTGCATCTGGTGGCCCTGAAACAGCCCGCAGCCG	60	
DB	22	CATATCTCATCAGGACTTTCAGCCAGTGTCTCCACTGGTGGCATGGAACACCCCGCTGTCT	81	
QY	61	GGCGCATGCGAGGACATCCCGGGAGCGGAGCTTCACGTGTTCACAGAGCGCGCGCCCGCG	120	
DB	82	GGAGGCATGCGTGTATCCGTGTAGCAGAGATTCCAGTGTTCACGCAAGCCCGAGCCCGTG	141	
QY	121	GGCTGGCGCGACCTTCGGGCGTTCCTGTCTGCTCGGGCTCGAGGACCTCTACAGCATC	180	
DB	142	GGCTGTGGGACACTTCGGGCTTCTGTCTCTTAGGCTGACAGATCTCTATAGCATC	201	
QY	181	GTGCGCCCGCCGACCCGACCGGAGTCCCGTGTCTCAACTCAGGGAACAGAGTGTCTTTC	240	
DB	202	GTGCGCCGTGTGACCCCGGGGTCTGTGCCCATCGTCAACTGAAGGAACAGAGTGTCTATCT	261	
QY	241	CCAGCTGGGAGGCTTATTCTCGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGGCATC	300	
DB	262	CCAGCTGGGACTCCCTGTTTCTGGCTCCAGGGTCAACTGCAACCCGGGGCCCGGCATC	321	

Db	322	TTTTTTTGTACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCGAGAAAGCGTATGG	381
Qy	361	CACGGCTCCGACCCAGCGGGCCGCCCTGACCGACAGCTACTGCGGAGACGTGGCGGACG	420
Db	382	CACGGCTCGACCCCGAGTGGCGGAGGCTGATGGAGTTACTGTGAGACATGGCGA	441
Qy	421	GAGCCCCCGGCGCCACCGGGCAGGCGTCGTGCTGCTGGCGGCGAGGCTGCTGAGCAG	480
Db	442	GAAGTACTGGGGCTACAGGTCAGGCTCCTCCCTGCTGTCAGGCAGGCTCCTGGACAG	501
Qy	481	GAGCGCGGAGCTGGCGCCACGCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC	540
Db	502	AAAGCTCGAGCTGCCACAAACAGCTACATCGTCTGCTGCTGCTGCTGCTGCTGCTG	561
Qy	541	TCCTTCTCCAA	551
Db	562	TCCTTCTCCAA	572

Search completed: March 29, 2004, 13:41:22
Job time : 59.1416 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 13:38:55 ; Search time 218.15 Seconds
(without alignments) 9471.775 Million cell updates/sec

Title: US-09-938-391-3

```

Perfect score: 555
Sequence: 1 cacaccacacggaacttcca.....tgacctctcttccaagtag 555

```

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2458946 pags. 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

Database : Published Applications NA:★

Database:

1:	/cgn2_6/pdata1/2/pubpna/us07_PUBCOMB.seq.*
2:	/cgn2_6/pdata1/2/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/pdata1/2/pubpna/us06_NEW_PUB.seq.*
4:	/cgn2_6/pdata1/2/pubpna/us06_PUBCOMB.seq.*
5:	/cgn2_6/pdata1/2/pubpna/us07_NEW_PUB.seq.*
6:	/cgn2_6/pdata1/2/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/pdata1/2/pubpna/us08_NEW_PUB.seq.*
8:	/cgn2_6/pdata1/2/pubpna/us08_PUBCOMB.seq.*
9:	/cgn2_6/pdata1/2/pubpna/us09A_PUBCOMB.seq.*
10:	/cgn2_6/pdata1/2/pubpna/us09B_PUBCOMB.seq.*
11:	/cgn2_6/pdata1/2/pubpna/us09C_PUBCOMB.seq.*
12:	/cgn2_6/pdata1/2/pubpna/us09_NEW_PUB.seq.*
13:	/cgn2_6/pdata1/2/pubpna/us10A_PUBCOMB.seq.*
14:	/cgn2_6/pdata1/2/pubpna/us10B_PUBCOMB.seq.*
15:	/cgn2_6/pdata1/2/pubpna/us10C_PUBCOMB.seq.*
16:	/cgn2_6/pdata1/2/pubpna/us10_NEW_PUB.seq.*
17:	/cgn2_6/pdata1/2/pubpna/us60_NEW_PUB.seq.*
18:	/cgn2_6/pdata1/2/pubpna/us60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	%			Length	DB	ID	Description
	Score	Match	Query				
1	555	100.0	555	10	US-09-938-391-3	Sequence 3, Appli	
2	555	100.0	829	10	US-09-938-391-1	Sequence 1, Appli	
3	550.4	99.2	552	14	US-10-131-241-50	Sequence 50, Appli	
4	550.4	99.2	552	14	US-10-952-418-34	Sequence 34, Appli	
5	418	75.3	546	14	US-10-042-347-4	Sequence 4, Appli	
6	418	75.3	549	14	US-10-131-241-53	Sequence 53, Appli	
7	418	75.3	549	14	US-10-952-418-3	Sequence 3, Appli	
8	418	75.3	552	9	US-09-873-676-30	Sequence 30, Appli	
9	418	75.3	3394	9	US-09-880-107-2178	Sequence 2178, Ap	
10	418	75.3	4551	14	US-10-060-036-144	Sequence 144, App	
11	418	75.3	4875	15	US-10-264-049-835	Sequence 835, App	
12	417	75.1	551	13	US-10-080-791-2	Sequence 2, Appli	
13	415.6	74.9	632	14	US-10-131-241-51	Sequence 51, Appli	
14	411.6	74.2	555	12	US-10-101-173-161	Sequence 161, App	
15	410.8	74.0	534	14	US-10-042-347-6	Sequence 6, Appli	

ALIGNMENTS

RESULT 1

US-09-938-391-3
; Sequence 3, Application US/09938391
; Publication No. US20030158099A1

GENERAL INFORMATION:

APPLICANT: Tong, et al.

1. TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING

TITLE OF INVENTION:	DISORDERS INVOLVING ANGIOGENESIS

FILE REFERENCE: PC10790A

; CURRENT APPLICATION NUMBER: US/09/938,391

;
;
; CURRENT FILING DATE: 2001-08-24

; NUMBER OF SEQ ID NOS: 4

```

; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 3

```

; LENGTH: 555
; TYPE: DNA

```

TYPE: DNA
ORGANISM:

US-09-938-391-3
ORGANISM: CANINE ENDOSTIAIN NUCLEOIDE SEQUENCE

0
9
8
7
6
5
4
3
2
1
0[illegible]

	QY	361	CACGGCTCCGACC	CGGCGGCCGCCTTGACCGACAGTACTTCGAGACAGTGGCGGACG	420
	Db	499	CACGGCTCCGACC	CGGCGGCCGCCTTGACCGACAGTACTTCGAGACAGTGGCGGACG	558
	QY	421	GAGCCCCCGCGC	ACCCGCGGCAGGGCTGTGCTCTCTGCGCGGCAGGGTGTCTGAGCAG	480
	Db	559	GAGCCCCCGCGC	ACCCGCGGCAGGGCTGTGCTCTCTGCGCGGCAGGGTGTCTGAGCAG	618
	QY	481	GAGCGCGAGCTG	CGCCACAGCCTTCGTGGTGTCTGATCGAGAACAGCGTCA TGACC	540
	Db	619	GAGCGCGAGCTG	CGCCACAGCCTTCGTGGTGTCTGATCGAGAACAGCGTCA TGACC	678
	QY	541	TCCTTCTCCAAGTAG	555	
	Db	679	TCCTTCTCCAAGTAG	693	
 RESULT 3					
US-10-131-241-50					
; Sequence 50, Application US/10131241					
; Publication No. US20030012792A1					
; GENERAL INFORMATION:					
; APPLICANT: Holaday, John W.					
; APPLICANT: Fortier, Anne H.					
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell					
; FILE REFERENCE: 05213-0344 43170-271565					
; CURRENT APPLICATION NUMBER: US/10/131,241					
; PRIOR FILING DATE: 2002-07-22					
; PRIOR APPLICATION NUMBER: US 09/413,049					
; PRIOR FILING DATE: 1999-10-06					
; PRIOR APPLICATION NUMBER: US 09/316,802					
; PRIOR FILING DATE: 1999-05-21					
; PRIOR APPLICATION NUMBER: US 60/086,586					
; PRIOR FILING DATE: 1998-05-22					
; NUMBER OF SEQ ID NOS: 65					
; SOFTWARE: PatentIn version 3.1					
; SEQ ID NO 50					
; LENGTH: 552					
; TYPE: DNA					
; ORGANISM: Canine sp.					
US-10-131-241-50					

[illegible]

Matches	466;	Conservative	0;	Mismatches	80;	Indels	0;	Gaps
1	QY	CACCCAC	CAGGACTTCCAGCTGGTGC	ACCTGGTGGCCCTGAA	CAGCCG	CAGCG	60	
1	Db	CACAGCAC	CGGACTTCCAGCCGGTGC	CTCCACCTGGTTGGCTCA	CAGCCCCCTG	CA	60	
61	QY	GGCGCAT	CGGAGCATCCGGGAGCGGACTT	CCAGTGTCTTCAG	CAGCGCGCGCCCG	C	120	
61	Db	GGCGCAT	CGGGGACATCCGGGGCCGACTT	CCAGTGTCTTCAG	CAGCGCGCGCGCCCG	C	120	
121	QY	GGGTGCG	CGGACCTTCGGGCGCTTCCTG	TGCTGGCGCTCAGAGCT	TACAGCATC		180	
121	Db	GGGTGCG	GGGACCTTCGGCGCTTCTGT	CTCGCGCTCAGAGCT	TACAGCATC		180	
181	QY	GTGCGCG	CGCGACCGCAC	CCGGGTCGCTGTC	CAACTCAGGGA	CAGGTGCTTTC	240	
181	Db	GTGCGCG	CTGCCACCGCG	CAGCGGTGCCATCGT	CAACTCAGGGA	CAGGTGCTTTC	240	
241	QY	CCAGCTGGG	GAGCCCTTATTC	TCGGGCTCCGAGGGCCAG	CTCAAGCCCGGGGCCCG	CATC	300	
241	Db	CCAGCTGGG	GAGCTGTCTTC	TCAGGCTCAGGGTCGG	CTCAAGCCCGGGGCCCG	CATC	300	

Query Match	75.3%	Score 418;	DB 9;	Length 552;
Best Local Similarity	85.3%;	Pred. No. 2.1e-86;		
Matches 466;	Conservative 0;	Mismatches 80;	Indels 0;	Gaps 0;
Qy	1	CACACCACACAGACTTCACGCTGCTGTCACCTGTGTCCTGAACAGCCCGCAGCCG	60	
Db	1	CACAGCACCGGACTTCACGCGGTGCTCCACCTGGTTGGGCTCAACAGCCCCCTGTCA	60	
Qy	61	GGCGGATCCGAGGATCCGGGGAGCGGACTTCACGTGCTTCCAGCAGCGCGCGCCGG	120	
Db	61	GGCGGATCCGGGGGATCCGGGGGCCGACTTCACGTGCTTCAGGAGCGCGCGGCCGGT	120	
Qy	121	GGGCTGGCGGCACTTCCGGGCTTCCTGTCGTCGGGCTGCAGACCTCTACAGCATC	180	
Db	121	GGGCTGGCGGGCACTTCCGGGCTTCTCTGCTGGCGCTGCAGGACTGTACAGCATC	180	
Qy	181	GTGCGCGGGCCGACCGCACCGGGTGCCCTCGTCAACCTCAGGACGAGTGTCTTC	240	
Db	181	GTGCGCGGTGCCACCGCGCACCGGTGCCCATCGTCAACCTCAAGACAGCTGTCTTT	240	
Qy	241	CCGAGTGGAGGCGCTTATTCGGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC	300	
Db	241	CCGAGTGGGAGGCTGTGTTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGGACGCACTC	300	
Qy	301	TTCTCTTTTCGAGCGGACAGATGTCCTGTCAGCACACCCCGCTGGCCCCGAGAGAGCGTGTGG	360	
Db	301	TTCTCTTTTGA CGGCAAGGAGCTCTTAGGCACACCCACCTGGCCCCCAGAGAGCGTGTGG	360	
Qy	361	CACGGCTCCGACCCGAGGGGGCGCGCTGTACCGACAGCTACTGCAGAGCGTGGCGGAGC	420	

Query Match	75.3%;	Score 418;	DB 9;	Length 3394;
Best Local Similarity	85.3%;	Pred. No. 1.6e-86;		
Matches 466;	Conservative 0;	Mismatches 80;	Indels 0;	Gaps 0;
QY 1	CACACCACAGGACTTCAGAGCTGGTCTGCACCTGGTGGCCCTGAAACGCCCGCAGCCG 60			
DB 1504	CACAGCCACCGCAGATTCCAGCGCGTGTCTCACTGGTTCGGTTCACAGCCCTCTGCA 1563			
QY 61	GGCGGCATGCGAGGCATCCGGGAGCGGGAATTCCAGTGTCTTCAGCAGGCGCGCGCCGCG 120			
DB 1564	GGCGGCATGCGGGGCATCCGCGGGCCGCACTTCCAGTGTCTTCAGCAGGCGCGGCGCGTG 1623			
QY 121	GGGCTGCGCGCACTTCCGGGCGCTTCCCTGTGTGTGGCGGTGCAGGACCTCTACAGCATC 180			
DB 1624	GGGCTGCGGGGCACTTCCGCGCGCTTCTCTCTCGCGCCTGCAGGACCTGTACAGCATC 1683			
QY 181	GTGCGCGCGCCGACCGCA CGGGGTGCCCGTGGTCAACTCAGGACGAGGTCTCTTTC 240			
DB 1684	GTGCGCGGTGCCGACCGCAGCGCGTGTCCCATGTCTCAACTCAGGACGAGCTGTCTTT 1743			
QY 241	CCCAGCTGGGAGGCTTATTCTCTGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCGCAATC 300			
DB 1744	CCCAGCTGGGAGGCTGTGTTCTCAGGCTCTGAGGTCCTGAAGCCCGGGGCAACGCATC 1803			
QY 301	TTCTCTTTCAGCGCAGAGATGTCTCTGAGCACCCTCGCTGGCCCCCGAAGAGCGCTGGG 360			
DB 1804	TTCTCTTTGACGGCAAGGACGTCTCTGAGGCACCCCACTGGCCCCCAGAGAGCGTGTGG 1863			
QY 361	CACGGCTCCGACCCACGCGGCGCGCTGACCCACACAGCTACTCGCAGACGTGGCGGACG 420			


```

; Sequence 51, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holladay, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Murinae sp.
; US-10-131-241-51

Query Match          74.9%; Score 415.6; DB 14; Length 632;
Best Local Similarity 82.8%; Pred. No. 7.2e-86;
Matches 523; Conservative 0; Mismatches 29; Indels 80; Gaps 1;

QY      1  CACACCCACGAGACTTCAGCTGGTGTGTCACCTGTGTGCCCTTGAAACAGCCGCGACCG 60
DB      1  CATACTCATCAGGACTTTCAGCCAGTGTCTCCACTGTGTGCACTGAACACCCCTCTGTCT 60

QY      61  GCGCGCATCGAGGGCATCCGGGAGCGGACTTCCAGTGTCTTCAGCAGGCGCGCGCGCG 120
DB      61  GAGGCGATCGCTGGTATTCGTGGAGCAGATTTCCAGTGTCTTCAGCAAGCCCGAGCGGTG 120

QY      121  GGCTCGCGCGCACCTTCGCGGGCTTCCTGTCTGTGCGGCG----- 160
DB      121  GGGCTGTGCGGCACCTTCGCGGGCTTTCTGTCTCTAGGCTGCAGGATCTTATGCAATC 180

QY      161  ----- 160
DB      181  GTCCGCGTGTCTACCGGGGTCTGTGCCCATCGTCAACCTGAAGCAGAGGTGCTATCT 240

QY      161  TGCGAGCACTCTACAGATCGTTCGCGCGCGCGCAGCCGCAACGGGGTCCCGTGTCTCAACC 220
DB      241  TGCGAGCACTCTACAGATCGTTCGCGCGCGCGCAGCCGCAACGGGGTCCCGTGTCTCAACC 300

QY      221  TCAGGACGAGGTGTCTTCCCGCAGCTGGAGAGCCCTTATTCTCGGGCTCCGAGGGCCAGC 280
DB      301  TCAGGACGAGGTGTCTTCCCGCAGCTGGAGAGCCCTTATTCTCGGGCTCCGAGGGCCAGC 360

QY      281  TGAAGCCCGGGGCCCGCATCTTCTTTTCGACGGCAGAGATGTCTTCAGCACCCTCGCT 340
DB      361  TGAAGCCCGGGGCCCGCATCTTCTTTTCGACGGCAGAGATGTCTTCAGCACCCTCGCT 420

QY      341  GGCCCGGAGAGCGTGTGTCAGCGGTCCCGACCCCGCGCGCGCGCTGCACCCAGCAGCT 400
DB      421  GGCCCGGAGAGCGTGTGTCAGCGGTCCCGACCCCGCGCGCGCGCTGCACCCAGCAGCT 480

QY      401  ACTCCGAGACGTGGCGGACGGAGCCCGCGCGCACCCGGGACGGCGTGTGTGTGTGTGG 460
DB      481  ACTCGAGACGTGGCGGACGGAGCCCGCGCGCACCCGGGACGGCGTGTGTGTGTGTGG 540

QY      461  CGGCGAGGTGTGGAGCAGAGGCGCGCAGCTGCCCGCACCGCTTCTGTGTGTGTGTGCA 520
DB      541  CGGCGAGGTGTGGAGCAGAGGCGCGCAGCTGCCCGCACCGCTTCTGTGTGTGTGTGCA 600

QY      521  TCAGAGAACAGCGTCATGACCTCTTCTCCAG 552
DB      601  TCAGAGAACAGCGTCATGACCTCTTCTCCAG 632

```

```

QY 541 TCTTTC 546
DB 3052 GCCTCC 3057

RESULT 12
US-10-080-797-2
; Sequence 2, Application US/10080797
; Publication No. US20020183253A1
; GENERAL INFORMATION:
; APPLICANT: Campochiaro, Peter A.
; APPLICANT: Dixon, Katharine H.
; APPLICANT: Brazzell, Romulus K.
; TITLE OF INVENTION: METHOD FOR TREATING OCULAR
; TITLE OF INVENTION: NEOVASCULARIZATION
; FILE REFERENCE: 4-31881A
; CURRENT APPLICATION NUMBER: US/10/080,797
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Human
US-10-080-797-2

Query Match 75.1%; Score 417; DB 13; Length 551;
Best Local Similarity 85.3%; Pred. No. 3.5e-86;
Matches 465; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 2 ACACCCACCCAGGACTTCCAGCTGGTGTGACACTGGTGGGCCCTGAACAGCCCGCAGCCGG 61
DB 1 ACAGCCACCGGACTTCCAGCGGCTGTCTCACTGGTGGCTCAACAGCCCCCTGTGAG 60
QY 62 GCGGCATGCGAGGCATCCGCGGAGCGGACTTCCAGTGTCTCCAGCAGGCGCGCGCCGG 121
DB 61 GCGGCATGCGGCGCATCCGCGGCGCGGACTTCCAGTGTCTTCCAGCAGGCGCGCGCGTGG 120
QY 122 GCGTGGCGGACACTTCCGGGCTTCTGTGTCGCGGCTGCGAGGACCTCTACAGCATCG 181
DB 121 GCGTGGCGGACACTTCCGGGCTTCTGTGTCGCGGCTGCGAGGACCTGTACAGCATCG 180
QY 182 TCGCGCGCGCGACCGCACCGGCGGTGCGCGTCTGTCACCTCAGGACGAGGTGCTCTTCC 241
DB 181 TCGCGCGGTGCGCACCGCGCGCGTGGCCCATCGTCAACTCAAGGACGAGCTGTGTTTC 240
QY 242 CCAGCTGGGAGGCTTATCTCGGCTCCGAGGCGCAGCTGAAGCCCGGGCCCGCATCT 301
DB 241 CCAGCTGGGAGGCTCTGTTCTCAGCTCTGAGGCTCGCTGAAGCCCGGGCAGCATCT 300
QY 302 TCTCTTTGACGGCAGAGATGCTGTGACGACCCCGGCTGGCCCCCGGAGAGAGCGTGTGGC 361
DB 301 TCTCTTTGACGGCAAGGACGCTCTGAGGACCCCGCATCTGGCCCCCAGAGAGCGTGTGGC 360
QY 362 AGGGCTCCGACCCAGCGGCGCGCTGACCCGACAGCTACTCCGAGAGCGTGGCGGACGG 421
DB 361 ATGGCTTCGACCCCAACGGGCGGAGGCTGACCCGAGAGCTACTGTGAGAGCTGGCGGACGG 420
QY 422 AGGCCCCCGGCGCCACCGGCGGAGGCGTGTGCTGTGGCGGCGAGGCTGCTGGAGCAGG 481
DB 421 AGGCTCCCTCGGCCACCGGCGGAGGCGTCTCGCTGTGGGGGCGAGGCTCTCTGGGGCAGA 480
QY 482 AGGCGCGAGCTCGGCCGACCGCTCTCGTGGTGTCTGCAATCGAGAAACAGCGTCATGACCT 541
DB 481 GTGCGCGGAGCTGCCATCAGCGCTACATCGTGCTCTGCAATTGAGAACAGCTTCATGACTG 540
QY 542 CTTTC 546
DB 541 CTTCC 545

```

RESULT 13
US-10-131-241-51

Matches 462; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 1 CACACCCACGAGGACTTCCAGCTGTGTCTCACACTGTGTGCTTGAACAGCCCGCAGCG 60
Db 4 CACAGCCACGCGACTTCCAGCGGTGCTCCACCTGTTGCGCTCAACAGCCCGCTGTCA 63
QY 61 GCGGCGATGCGAGGATTCGGGAGCGGATTCAGTGTTCCTCAGCAGGCGCGCGCG 120
Db 64 GCGGCGATGCGGCGGATTCGGGAGCGGATTCAGTGTTCCTCAGCAGGCGCGCGCG 123
QY 121 GGGTGGCGCGACCTTCCGGGCGCTTCTGTGTGTGCGGCTGAGGACCTTACAGCATC 180
Db 124 GGGTGGCGCGACCTTCCGGGCGCTTCTGTGTGTGCGGCTGAGGACCTTACAGCATC 183
QY 181 GTGCGCGCGCGGACCGACCGGGGTGCCGTGCTCAACTCAGGCGACGAGTGTCTTTC 240
Db 184 GTGCGCGGTGCGGACCGCGGAGCGGTGCCATCTCAACTCAGGCGAGTGTCTTTC 243
QY 241 CCCAGCTGGGAGGCTTATTTCTCGGGCTCCAGGCGGACGTGAAGCCCGGGCGCGCATC 300
Db 244 CCCAGCTGGGAGGCTTGTGTTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGCGCGCATC 303
QY 301 TTCTCTTTTCGAGCGGACAGAGATGCTTCAGCACCCCGCTGGCCCGGAGAGCGTGTGG 360
Db 304 TTCTCTTTTAAGCGAAGGACGCTCTGACCCACCCACTTGGCCCCCAGAAAGCGTGTGG 363
QY 361 CACGGCTCCGACCCAGCGGCGCGCGCTGACCGACGCTACTGCGAGACGTGGCGGACG 420
Db 364 CATGGCTCGGACCCCAACGGGCGCGAGGCTACCGGAGAGCTACTGTGAGACGTGGCGGACG 423
QY 421 GAGCGCCCGGCGGACCGCGGCGAGGCGTCTGCTGCTGCGCGGCGAGGCTGCTGGAGCAG 480
Db 424 GAGGCTCCCTCGGCGACGCGGCGAGGCTACTGCTGCTGGGGGCGAGGCTCTGGGGGCGAG 483
QY 481 GAGCGCGGAGCTGCCCGCACGCTTCTGTGTGTGCTGTGATCGAGAACGCGTCAAGCC 540
Db 484 AGTGGCGGAGCTGCCATCAGCGCTACATCGTGTATGATTGAGAACAGCTTCATGACT 543
QY 541 TCCTTC 546
Db 544 GCCTCC 549

RESULT 15
US-10-042-347-6
; Sequence 6, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide 1
; FILE OF INVENTION: Thereof
; FILE REFERENCE: 05213-0880 (43170-249874)
; CURRENT APPLICATION NUMBER: US/10/042,347
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 09/315,689
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 60/106,343
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 09/154,302
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: US 08/740,168
; PRIOR FILING DATE: 1996-10-22
; PRIOR APPLICATION NUMBER: US 60/005,835
; PRIOR FILING DATE: 1995-10-23
; PRIOR APPLICATION NUMBER: US 60/023,070
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: US 60/026,263
; PRIOR FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 534

```

; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6
; LENGTH: 534

```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-347-6

Query Match      74.0%; Score 410.8; DB 14; Length 534;
Best Local Similarity 85.6%; Pred. No. 9.2e-85;
Matches 457; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 13 GACTTCAGTGGTGTGACCTGGTGGCCCTGAACAGCCGCGGCGGCGCATGCGA 72
Db 1 GACTTCAGCGGTGTCTCACTGGTGGCTCAACAGCCCCCTGTACAGGGCGCATGGG 60

QY 73 GGCATCCGGGGAGCGAATTCAGTCTTCACAGCAGCGCGCGCGCGGGCTGGCCGGC 132
Db 61 GGCATCCGGGGCGGACATTCAGTCTTCACAGCAGCGCGCGCGCGGGCTGGCGGGC 120

QY 133 ACCTTCGGGSCCTTCCTGTCTCGCGGCTGCAGGACCTCTACAGCATGCTGCGCGCGCC 192
Db 121 ACCTTCGGGCCCTTCCTGTCTCGCGCTTCAGGACCTGTACAGCATGCTGCGCGGTGCC 180

QY 193 GACCGCACCGGGGTGCCCGTCTCAACCTCAAGGACGAGGTCTTCCCGCAGCTGGGAG 252
Db 181 GACCGCGCAGCGGTGCCCATCGTCAACCTCAAGGACGAGCTGTCTTCCAGCTGGGAG 240

QY 253 GCCTTATCTCGGGCTCCGAGGCGCAGCTGAAGCCCGGGGCGCGCATCTTCTCTTTGAC 312
Db 241 GCTCTGTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGGCGCAGCATCTTCTCTTTGAC 300

QY 313 GGCAGAGATGTCTTGACACCCCGCTTGGCCCCCGAAGAGCGTGTGGCAGCGCTCCGAC 372
Db 301 GGCAGAGACGTCTTGAGGCACCCCACTTGGCCCCAGAGAGCGTGTGGCATGGCTCGGAC 360

QY 373 CCCAGCGGGCGCCCTGACCGACAGCTACTGCGAGACCTGGGGACGAGAGGCCCGGGCG 432
Db 361 CCCAACGGGGCGCAGGCTGACCGAGAGCTACTGTGAGACGTGGCGGACGAGGCTCCCTCG 420

QY 433 GCCACCGGCGAGGCGTCTGCTGCGCGGCGAGGCTGCTGGAGCAGGAGGCCCGCGAGC 492
Db 421 GCCACGGGCGAGGCGCTCTGCTGCTGGGGGCGAGGCTCTTGGGGCAGAGTSCCGCGAGC 480

QY 493 TGCCGCCACGCGCTTCGTGGGTGCTCTGCATCGAGAACAGCGTCAATGACCTCCTTC 546
Db 481 TGCCATCAGCGCTACATCGTGTCTCTGCAITTGAGACAGCTTTCATGACTGCCTCC 534
```

Search completed: March 29, 2004, 16:40:38
Job time : 231.15 secs

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 11:07:01 ; Search time 1704.7 Seconds
(without alignments)
9722.244 Million cell updates/sec

Title: US-09-938-391-3

Perfect score: 555

Sequence: 1 cacaccaccaggacttcca.....tgacctctctccaaagtag 555

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_estli:*

10: gb_est2:*

11: gb_estc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_hum:*

19: em_gss_hum:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_pro:*

26: em_gss_pig:*

27: em_gss_vrt:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	439.8	79.2	551	10	BF074459 221883 MA
2	418	75.3	707	10	BE908201 601500458
3	418	75.3	757	10	BE908253 601502237
4	418	75.3	881	14	CD105862 AGENCOURT

5	418	75.3	4230	11	BC063833	BC063833 Homo sapi
6	410.8	74.0	504	14	CF789984	CF789984 867322 MA
7	394.2	71.0	944	13	EU859398	EU859398 AGENCOURT
8	391.8	70.6	657	14	CB444165	CB444165 635295 MA
9	384.4	69.3	703	13	BU615520	BU615520 UI-H-FGO-
10	384.2	69.2	929	13	BQ672290	BQ672290 AGENCOURT
11	377	67.9	683	12	BM683067	BM683067 UI-E-BOL-
12	376.6	67.9	682	9	AW089583	AW089583 x20f03.x
13	375.8	67.7	843	10	BF385854	BF385854 602046021
14	375.8	67.7	874	12	BI412588	BI412588 602990468
15	375.8	67.7	4192	11	BC082931	BC082931 Mus muscu
16	374.8	67.5	715	14	CB596713	CB596713 AGENCOURT
17	373.2	67.2	735	14	CF728236	CF728236 UI-M-HBO-
18	371.6	67.0	832	12	BG387051	BG387051 602454749
19	370.2	66.7	979	13	BQ673186	BQ673186 AGENCOURT
20	369.6	66.6	897	12	BI080524	BI080524 602877005
21	368.6	66.4	611	10	AW911243	AW911243 ut83h10.Y
22	367.2	66.2	884	12	BI161007	BI161007 602865213
23	366.6	66.1	1093	13	BQ723254	BQ723254 AGENCOURT
24	364.2	65.6	723	12	BI247582	BI247582 602960041
25	363.8	65.5	753	12	BI904605	BI904605 603168411
26	362.6	65.3	782	9	AI326391	AI326391 mm18f09.x
27	360.2	64.9	668	14	CF724654	CF724654 UI-M-GZO-
28	357.4	64.4	618	9	AV696242	AV696242 AV696242
29	350.6	63.2	664	13	BU632049	BU632049 UI-H-PE1-
30	350.6	63.2	715	9	AU125614	AU125614 AU125614
31	350.6	63.2	947	13	BU556872	BU556872 AGENCOURT
32	348.6	62.8	720	12	BI147444	BI147444 602914008
33	348.6	62.8	915	10	BF166139	BF166139 601776586
34	344.6	62.1	652	13	BU352506	BU352506 603527982
35	344.2	62.0	650	9	AI858615	AI858615 w140f01.x
36	342.6	61.7	639	12	BM988137	BM988137 UI-H-DT1-
37	341.4	61.5	614	13	BU455935	BU455935 60367327
38	341.4	61.4	634	13	BQ772348	BQ772348 UI-H-EZ1-
39	340.4	61.3	587	9	AA288198	AA288198 vb15b10.r
40	340	61.3	660	10	AW192502	AW192502 x145e10.x
41	339.2	61.1	690	12	BI219399	BI219399 602936756
42	336.2	60.6	795	14	CB601615	CB601615 AGENCOURT
43	335.2	60.4	634	9	AI970297	AI970297 wr09c02.x
44	334.8	60.3	706	13	BU439577	BU439577 603208252
45	333.8	60.1	745	10	AW243446	AW243446 xm95d11.x

ALIGNMENTS

RESULT 1

BF074459

LOCUS

221883 MARC 2BOV Bos taurus cDNA 5', mRNA

BF074459

DEFINITION

BF074459.1 GI:10867970

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Bos taurus (cow)

REFERENCE

AUTHORS

Smith, I.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,

Casas, E., Wray, J.E., White, J., Cho, J., Fahrkrug, S.C.,

Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,

Chitko-McKown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F.,

Quackenbush, J. and Keefe, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

JOURNAL

MEDLINE

11280013

PUBMED

11282978

COMMENT

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers

FORWARD: AGAAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 81 row: F column: 9

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

1. 551

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/tissue_type="pooled"

/lab_host="DH10B"

/clone_lib="MARC 2BOV"

/note="Vector: pCMV SPOR6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semiteadonous muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

FEATURES

source

ORIGIN

Query Match 79.2%; Score 439.8; DB 10; Length 551;
Best Local Similarity 89.3%; Pred. No. 3.8e-53;
Matches 474; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 25 GTGCTGACCTGCTGGCCCTTGAACAGCCCGGCGGCGATGAGAGCATCCGGGA 84

Db 1 GTGCTGACCTGCTGGCCCTTGAACAGCCCGGCGGCGATGAGAGCATCCGGGC 60

QY 85 GCGGACTTCCAGTCTTCCAGAGCGCGCGCGGCGGCTGCGCGGACCTTCGGGCG 144

Db 61 GCGGACTTCCAGTCTTCCAGAGCGCGCGCGGCGGCTGCGCGGACCTTCGGGCG 120

QY 145 TTCTGTCGTCGCGGTCGAGGACTCTACAGCATGCTGCGCGCGCGGCGGCGGCGG 204

Db 121 TTCTGTCGTCGCGGTCGAGGACTCTACAGCATGCTGCGCGCGCGGCGGCGGCGG 180

QY 205 GTGCCGCTGCTCAACCTCAGGAGAGGCTGCTTCCAGCTGGGAGGCTTATTTCTCG 264

Db 181 CTGCCGCTGCTCAACCTCAGGAGAGGCTGCTTCCAGCTGGGAGGCTTATTTCTCA 240

QY 265 GCTCCGAGGCGCAGCTGAACCCGGGCGCGGCTTCTTCTTCGACGCGAGATGTC 324

Db 241 GCTCCGAGGCGCAGCTGAACCCGGGCGCGGCTTCTTCTTCGACGCGAGATGTC 300

QY 325 CTGACGACCCCGCTGCGCGCGGCGGAGGCTGTGGCAACGCTCCGACCCCGGCGCG 384

Db 301 CTGACGATCCCACTGCGCGCGGCGGAGGCTGTGGCAACGCTCCGACCCCGGCGCG 360

QY 385 CGCTTGACCGACGCTACTGCGAGAGCTGTGGCGGACGAGGCGCGCGCGCGCGGCGG 444

Db 361 CGCTTGACCGACGCTACTGCGAGAGCTGTGGCGGACGAGGCGCGCGCGCGCGGCGG 420

QY 445 GCGTCTGCTGCTGCGCGGCGGCTGCTGGAGCAGGAGCGCGGAGCTGCGCGCACGCG 504

Db 421 GCGTCTGCTGCTGCGCGGCGGCTGCTGGAGCAGGAGCGCGGAGCTGCGCGCACGCG 480

QY 505 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 555

Db 481 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531

RESULT 2

BE908201

LOCUS

601500458F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3902175 5',

DEFINITION

mRNA sequence.

ACCESSION

BE908201

VERSION

BE908201.1

GI:10402537

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9704 row: d column: 16

High quality sequence stop: 688.

Location/Qualifiers

1. .707

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3902175"

/tissue_type="epithelioid carcinoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 70"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.1 kb. Library constructed by Life

Technologies."

ORIGIN

Query Match 75.3%; Score 418; DB 10; Length 707;

Best Local Similarity 85.3%; Pred. No. 4.9e-50;

Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACAGACATTCACAGCTGGTGTGCACCTGGTGGCCCTGAAACAGCCCGGAGCGG 60

Db 11 CACAGCCACCGGACATTCACAGCTGGTGTGCACCTGGTGGCCCTGAAACAGCCCGGAGCGG 70

QY 61 GCGGCGCATGCGAGGCGATTCGGGGAGCGGACTTCCAGTGTTCAGAGCGGCGCGCGCG 120

Db 71 GCGGCGCATGCGGGGCGATTCGGGGAGCGGACTTCCAGTGTTCAGAGCGGCGCGCGCG 130

QY 121 GGGTGGCGGCGACCTTCGGGGCTTCCCTGTCGCGGCTGCGAGGACCTCTACAGCATC 180

Db 131 GGGTGGCGGCGACCTTCGGGGCTTCCCTGTCGCGGCTGCGAGGACCTCTACAGCATC 190

QY 181 GTGGCGCGCGCGGACCGGCGGAGCGGCTGCTCAACCTCAGGACGAGGTGCTCTTC 240

Db 191 GTGGCGCGCGGACCGGCGGAGCGGCTGCTCAACCTCAGGACGAGGTGCTCTTC 250

QY 241 CCCAGCTGGAGGCGCTTATTCCTGGGCTCGAGGCGGAGCTGAAGCCCGGGCGCGCATC 300

Db 251 CCCAGCTGGAGGCGCTTATTCCTGGGCTCGAGGCGGAGCTGAAGCCCGGGCGCGCATC 310

QY 301 TTCTCTTTGACGCGCAGAGATGCTCTGCGACACCCCGCTGGCCCGGCGGAGAGCGTGTGG 360

Db 311 TTCTCTTTGACGCGCAGAGATGCTCTGCGACACCCCGCTGGCCCGGCGGAGAGCGTGTGG 370

QY 361 CACGGCTCCGACCCCGGCGGCGGCGGCTGACGACGAGCTACTGCGAGAGCTGCGGCGGAG 420

Db 371 CATGGCTCGGACCCCGGCGGCGGCGGCTGACGAGAGCTACTGCGAGAGCTGCGGCGGAG 430

QY 421 GAGGCCCGGCGGCGGCGGCGGCGGCTGCTGCTGCTGGCGGCGGAGGTGCTGGAGCGAG 480

Db 431 GAGGCTCCCTCGGCGGCGGCGGCGGCGGCTGCTGCTGCTGGCGGCGGAGGTGCTGGAGCGAG 490

QY 481 GAGGCCCGGAGCTGCGGCGGCGGCGGCTGCTGCTGCTGCTGCGAGACGCGTCTGAGCC 540

Db 491 AGTGCCGCGAGCTGCCATCAGCGCTCATCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550

QY 541 TCCTTC 546
 Db 551 GCCTCC 556

RESULT 3

BE906253
 LOCUS 601502237F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3904208 5',
 DEFINITION linear EST 20-OCT-2000

ACCESSION BE906253
 VERSION BE906253.1 GI:10399595

KEYWORDS EST

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM9709 row: 1 column: 09

High quality sequence stop: 757.

Location/Qualifiers

1. 757

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="IMAGE:3904208"

/tissue_type="epitheloid carcinoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 70"

/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.1 Kb. Library constructed by Life

Technologies."

ORIGIN

Query Match 75.3%; Score 418; DB 10; Length 757;

Best Local Similarity 85.3%; Pred. No. 5e-50;

Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACGAGACTCCAGCTGGTGTGTCACCTGGTGGCCCTGAACAGCCCGCAGCG 60

Db 36 CACAGCCACCGCACTCCAGCCGCGTCCACCTGGTGGCCCTGAACAGCCCGCAGCG 95

QY 61 GCGCGCATGCGAGGATCCGCGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCG 120

Db 96 GCGCGCATGCGGCGATCCGCGGCGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCG 155

QY 121 GCGGTGCGCGGCACTTCCGCGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180

Db 156 GCGGTGCGCGGCACTTCCGCGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 215

QY 181 GTGCGCGCGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 240

Db 216 GTGCGCGCGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 275

QY 241 CCAGCTGGGAGGCGCTTATCTCGGCGCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCG 300

Db 276 CCAGCTGGGAGGCGCTTATCTCGGCGCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCG 335

QY 301 TTCTCTTTCAGCGGAGAGATGTCCTGTGAGCACCCCGCGCTGCGCGCGCGCGCGCG 360

Db 336 TTCTCTTTCAGCGGAGAGATGTCCTGTGAGCACCCCGCGCTGCGCGCGCGGCGG 395

QY 361 CACGCTCGGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420

Db 396 CATGCTCGGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 455

QY 421 GAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

Db 456 GAGGCTCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 515

QY 481 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 540

Db 516 AGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 575

QY 541 TCCTTC 546

Db 576 GCCTCC 581

CD105862 881 bp mRNA linear EST 15-MAY-2003

AGENCOURT 14021788 NIH MGC 179 Homo sapiens cDNA clone

IMAGE:30365831 5', mRNA sequence.

CD105862

CD105862.1 GI:30759036

KEYWORDS EST

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Michael Brownstein

cDNA Library Preparation: Invitrogen Corp

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: NDAM424 row: 9 column: 24

High quality sequence stop: 689.

Location/Qualifiers

1. 881

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="IMAGE:30365831"

/tissue_type="Pituitary"

/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"

/clone_lib="NIH MGC 179"

/notes="Organ: brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV

(destroyed); Site 2: NotI; Library is oligo-dT primed and

directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.1 kb. Library was

constructed by (Invitrogen). Note: this is a NIH_MGC

Library."

ORIGIN

Query Match 75.3%; Score 418; DB 14; Length 881;

Best Local Similarity 85.3%; Pred. No. 5.1e-50;

Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 CACACCCACGAGACTTCCAGCTGGTGTGTCACCTGGTGGCCCTGAACAGCCCGCAGCG 60

Db 292 CACAGCCACCGCACTTCCAGCCGCGTGTCCACCTGGTGGCCCTGAACAGCCCGCAGCG 351

QY 61 GCGCGCATGCGAGGATCCGCGGAGCGGACTTCCAGTGTCTCCAGCAGCGCGCGCGCG 120

352	GGCGGCATGCGGGGCATTCGCGGGCGCATTCAGTGCTTCCAGCAGGCGCGGGCGG	411
121	GGGCTGGCGCGACCTTCCGGGCGCTTCTGTCTCGCGCTCGAGGACCTCTACAGCATC	180
412	GGGCTGGCGGGACCTTCCGGCGCTTCTGTCTCGCGCTCGAGGACCTGTACAGCATC	471
181	GTGGCGCGCGCCAGCGCACCGGGGTGCCGTGCTCAACTCAGGCGCAGAGTGTCTTTC	240
472	GTGGCGCGTGCAGCCGCGCGACCGTGGCCATCGTCAACTCAGGCGCAGCTGCTGTT	531
241	CCAGCTGGGAGGCGCTTATCTCGGGCTCCGAGGGCGAGCTGAAGCCCGGGGCGCGATC	300
532	CCCAGCTGGGAGGCTCTGTGTCTCAGGCTCTGAGGCTCGCGCTGAAGCCCGGGGACGATC	591
301	TTCTCTTTTCGACGGCAGAGATGCTCTGCAGCACCCCGGCTGTGCCCCGGAAGAGCTGTGG	360
592	TTCTCTTTTCGAGCAAGAGCGTCTTGAGSCACCCCACTGGCCCCCAGAGAGCGGTGG	651
361	CAGGGCTCCGAGCCCGAGGGGCGCGCTGACCGACAGCTACTGCGAGAGCTGGGCGAGCG	420
652	CATGGCTTCGAGCCCCCAAGGGCGAGGCTGACCGAGAGCTACTGTGAGAGCGTGGCGGAG	711
421	GAGGCCCCCGCGCCACCGGGCAGGCGCTCGCTGCTGGCGGGCAGGCTGCTGGAGCAG	480
712	GAGGCTCCCTCGGCCACCGGCCAGGCGCTCTCTCGCTGCTGGGGGCGAGGCTCTGGGCGAG	771
481	GAGGCGCGAGTGGGCGCAGCGCTTGTGGTGCTCTGCATCGAGACAGCGTCAATGACC	540
772	AGTGGCGGAGTGGCATCAGCGCTCATCGTGCTCTGCATTTGAGAACAGCTTCATGACT	831
541	TCCTTC	546
832	GCCTCC	837

RESULT 5
BC063833
LOCUS BC063833
DEFINITION Homo sapiens cDNA clone IMAGE:61818; containing frame-shift
4230 bp mRNA linear HTC 09-DEC-2003

errors.
ACCESSION BC063833
VERSION BC063833.1 GI:39645297
KEYWORDS HTC.

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata;
Mammalia; Eutheria; Primates;
Cataarrhini; Hominoidea; Homo.

1 (Dasses 1 to 4230)
Srausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Shat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.K.,
Diatchenko,L., Marusina,K., Parnet,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,F., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Caranci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahey,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whitting,M., Madan,A.C., Young,A.C., Schenchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodrigues,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Batterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED
12477932

2 (Dasses 1 to 4230)
Srausberg,R.

REFERENCES
AUTHORS

TITLE	Direct Submission
JOURNAL	Submitted (08-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: gcapob-remail.nih.gov Tissue Procurement: Dr. James R. Lupski cDNA library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcsc.bc.ca
	Steven Chan, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chen, Readman Chiu, Chris Fiell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Ness, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saesedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: TRAK Plate: 141 Row: C Column: 6 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18765747 This clone has the following problem: frame shifted.
FEATURES	Location/Qualifiers 1..4230 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6181818" /tissue_type="Peripheral Nervous System, dorsal root ganglion" /clone_lib="Lupski_dorsal_root_ganglion" /lab_host="DH10B" /note="Vector: pCMV-SPORT6"
ORIGIN	Query Match 75.3%; Score 418; DB 11; Length 4230; Best Local Similarity 85.3%; Fred. No. 6.7e-50; Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY	1 CACACCCACGAGACTTCCAGCTGGTGTGTGACCTGGTGGCCCTGTGAACAGCCCGCAGCCG 60
Db	3495 CACAGCCACCGGACTTCCAGCGGTGTCTCCACTGGTGGCTCAACAGCCCTGTCA 3554
QY	61 GCGCGATGCGAGCATCCGGAGGAGGACTTCCAGTGTCTCCAGAGCGCGCGCGCG 120
Db	3555 GCGCGATGCGGGGCAATCCGGGGGCGCACTTCCAGTGTCTCCAGAGCGCGCGCGCGT 3614
QY	121 GGGTGGCGCGACCTTCCGGGCGCTTCTGTGTGTGGGGTGCAGGACCTCTACAGCATC 180
Db	3615 GGGTGGCGGGCACTTCCGGCGCCCTTCTGTCTTCCGGGCTGCAGGACCTGTACAGCATC 3674
QY	181 GTGCGCGCGCGCACCGAGCGCGGTGCCGTGTCACTCAAGTCCAGGACGAGGTCCTTC 240
Db	3675 GTGCGCGGTGCGACCGCGAGCCGTGCGCATGTCTAACTCANGAGCAGCTGTCTGTTT 3734
QY	241 CCCAGCTGGAGGACCTTATTCTCGGGCTCCAGGGCGCAGGTGAAGCCGGGCGCCGATC 300
Db	3735 CCCAGCTGGAGGAGCTCTGTCTTCAGGCTCTGAGGGTCCGTGAAGCCGGGCGCAGCATC 3794
QY	301 TTCTCTTTCCAGCGCAGATGCTCTCTGACGACACCCCGCTTGGCCCGAGAGCGTGTGG 360
Db	3795 TTCTCTTTGAGCGAAGGACGCTCTTGAGGACCCCGACCTGGCCCCCAGAGCGTGTGG 3854
QY	361 CAGGCTTCGACCCCGAGCGGGCGCCGCTTGACCGACAGTACTGTGAGAGCGTGGCGGACG 420
Db	3855 CATGGCTCGACCCCAACGGGCGCAGGTGACCGAGAGCTACTGTGAGAGCTGCGCGGCG 3914

181 GGGGGGGGAGGCGCTTGTCTGGGCTCTGAGGCGCAGCTGAAGCCCGCGCCGATCT 240
302 TCTCTTTTCAGCGGAGAGATGTCTGTGAGCACCCTGCTGCGCCCGGAGAGCGTGTGGC 361
241 TCTCTTTTCAGCGGAGAGATGTCTGTGAGCACCCTGCTGCGCCCGGAGAGCGTGTGGC 300
362 AGGCTCTCGACCCAGCGGCGCGCTGACCGACAGCTACTGCGAGACGTGGCGGACGG 421
301 AGGCTCTCGACCCAGCGGCGCGCTGACCGACAGCTACTGCGAGACGTGGCGGACGTG 360
422 AGGCTCTCGACCCAGCGGCGCGCTGACCGACAGCTACTGCGAGACGTGGCGGACGG 481
361 AGGCTCTCGACCCAGCGGCGCGCTGACCGACAGCTACTGCGAGACGTGGCGGACGG 420
482 AGGCTCTCGACCCAGCGGCGCGCTGACCGACAGCTACTGCGAGACGTGGCGGACGG 541
421 AGGCTCTCGACCCAGCGGCGCGCTGACCGACAGCTACTGCGAGACGTGGCGGACGG 480
542 CTTTCTTCCAAAGTAG 555
481 CCGCTTCCAAAGTAG 494

RESULT 7
BU859398/c
LOCUS
DEFINITION
AGENCOURT 10434549 NIH MGC 109 Homo sapiens cDNA clone
IMAGE:6650260 5', mRNA sequence.
BU859398
BU859398.1 GI:24044390
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BU859398 944 bp mRNA linear EST 16-OCT-2002
AGENCOURT 10434549 NIH MGC 109 Homo sapiens cDNA clone
IMAGE:6650260 5', mRNA sequence.
BU859398
BU859398.1 GI:24044390
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM2895 row: d column: 04
High quality sequence stop: 572.
Location/Qualifiers
1. .944
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6650260"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/notes="Organ: ovary; Vector: pORF7; Site: 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G) Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life technologies). Note: this is a
NIH_MGC Library."

FEATURES
source
1. .944
Location/Qualifiers

ORIGIN
Query Match 71.0%; Score 394.2; DB 13; Length 944;
Best Local Similarity 83.7%; Pred. No. 1.2e-46;
Matches 458; Conservative 0; Mismatches 88; Indels 1; Gaps 1;

421 GAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
3915 GAGGCTCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3974
481 GAGGCGCGGAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
3975 AGTCCGCGGAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4034
541 TCCTTC 546
4035 GCCTCC 4040

RESULT 6
CF789984
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CF789984 504 bp mRNA linear EST 21-OCT-2003
867322 MARCH 4PIG Sus scrofa cDNA 5', mRNA sequence.
CF789984
CF789984.1 GI:37794545
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 504)
Smith, T.P.L., Freking, B.A., Ford, J.J., Vallet, J.L., Wise, T.A.,
Noneman, D.J., Wray, J.E. and Keele, J.W.
Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim-alt option. Vector identified with
cross_match v0.990329.
Plate: TW8006 row: B column: 3
Seq primer: GTAATACGACTCTATAGG.
Location/Qualifiers
1. .504
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4PIG"
/notes="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."

FEATURES
source
1. .504
Location/Qualifiers

ORIGIN
Query Match 74.0%; Score 410.8; DB 14; Length 504;
Best Local Similarity 89.5%; Pred. No. 4.9e-49;
Matches 442; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
62 GCGGCGTCCGAGGATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 121
1 GCGGCGTCCGAGGATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
122 GGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 181
61 GGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
182 TCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
121 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
242 CCAGCTGGAGCGCTTATCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301

```
QY 1 CACACCACAGGAGCTTCCAGCTGGT-GCTGACCTGTGGCCCTGAAACCCCGGAGCC 59
DB 712 CAGGTCCACCGAGCTTCCAGCGGTGGCTCCACATGGTGGCTCAGACGCCCTTCTC 653
QY 60 GGGCGGATCGAGGAGCTCCGGGAGCGGACTTCCAGTCTCCAGCAGGCGCGCGGC 119
DB 652 AGGCGGATCGGGGATCCCGGGCCCGACCTTCCAGTCTCCAGCAGGCGCGCGGC 593
QY 120 GGGGCTGGCGGCACTTCCGGGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 179
DB 592 GGGGCTGGCGGCACTTCCGGGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 533
QY 180 GTGCGCGCGCGGCGAGCGGACCGGGGTGCGGTCTGTCTGTCTGTCTGTCTGTCT 239
DB 532 GTGCGCGCGCGGCGAGCGGACCGGGGTGCGGTCTGTCTGTCTGTCTGTCTGTCT 473
QY 240 CCCAGCTGGGAGGCTTATTTTCGGGCTCCGAGGCGGAGCTGAGCGGCGCGCAT 299
DB 472 TCCAGCTGGGAGGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 413
QY 300 CTTCCTTTTCGAGCGGAGAGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 359
DB 412 CTTCCTTTTCGAGCGGAGAGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 353
QY 360 GCACGCTTCGAGCGGAGGCGGCGGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 419
DB 352 GCATGGCTTCGAGCGGAGGCGGCGGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 293
QY 420 GGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 479
DB 292 GGAGGCTTCCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 233
QY 480 GGAGGCGCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 539
DB 232 GAGTGGCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 173
QY 540 CTCCTTC 546
DB 172 TGCTTC 166
```

```
RESULT 8
LOCUS CB444165 657 bp mRNA linear EST 25-MAR-2003
DEFINITION 695295 MRC 6BOV Bos taurus cDNA 3', mRNA sequence.
ACCESSION CB444165
VERSION CB444165.1 GI:292333914
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 657)
Smith, J.F.L., Roberts, A.J., Ehternkamp, S.E., Chitko-McKown, C.G.,
Wray, J.E. and Keele, J.W.
A second set of bovine ESTs from pooled-tissue normalized libraries
Unpublished (2003)
Contact: Smith JPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.
Plate: FQY8051 row: 1 column: 11
Seq primer: TAGAGGCGACAGTCGAGG.
Location/Qualifiers
1..657
/organism="Bos taurus"
/mol_type="mRNA"
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

```
FEATURES
source
1..657
/organism="Bos taurus"
/mol_type="mRNA"
```

```
/db xref="taxon:9913"
/tissue_type="pooled"
/lab host="DH10B"
/clone lib="MARC 6BOV"
/notes="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."
```

```
ORIGIN
Query Match 70.6% Score 391.8; DB 14; Length 657;
Best Local Similarity 88.9%; Pred. No. 2.5e-46;
Matches 423; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 80 GGGAGGCGGCTTCCAGTGTCTCCAGCAGCGCGCGGCTGCGCGGCTGCGCGGCTTCC 139
DB 657 GCGCGCGGCTTCCAGTGTCTCCAGCAGCGCGCGGCTGCGCGGCTGCGCGGCTTCC 598
QY 140 GGGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 199
DB 597 GCGGTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 538
QY 200 CCGGGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 259
DB 537 CCACTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 478
QY 260 TCTCGGCTTCCAGGCGGCTGAGCGCGGCGGCGGCGGCTTCTGTCTGTCTGTCTGTCTGT 319
DB 477 TCTCAGCTTCCAGGCGGCTGAGCGCGGCGGCGGCGGCTTCTGTCTGTCTGTCTGTCTGT 418
QY 320 ATGCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 379
DB 417 ATGCTCTTCCAGCTTCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 358
QY 380 GGGCGGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 439
DB 357 GCGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 298
QY 440 GCGAGGCGGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 499
DB 297 GCCAGGCGGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 238
QY 500 ACGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 555
DB 237 ACGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 182
```

```
RESULT 9
LOCUS BU615520 703 bp mRNA linear EST 23-SEP-2002
DEFINITION UI-H-FG0-bcp-p-10-0-UI-s1 NCI CGAP ENL 2 Homo sapiens cDNA clone
ACCESSION BU615520
VERSION BU615520.1 GI:23281735
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 703)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
```


Db 527 TTCTCTTTAACGGCAAGACGCTCTGAGGCCACCCACCTGGCCGCCAGAGAGCGTGTGG 586
 Qy 361 CACGGCTCCGACCCACCGAGCGGCGCGCTGACCGACAGTACTGCGACAGCTGGCGGACG 420
 Db 587 CATGGCTCGGACCCCAACAGCGGCGCAGGCTGACCGAGAGTACTGTGAGAGCTGGCGGACG 646
 Qy 421 GAGGCCCCGGCGCCACCGGCGGAGGCTGTGCTGCTGCTGGCGGCGAGG---CTCTGGAG 477
 Db 647 GAGGCTCCCTCGGCGACGGGCGCAGGCTCTCTCCCTGCTGGGCGGCGAGCTCCCTGGGCG 706
 Qy 478 CAGGAGCCCGGAGCTGCGCCACCGCTTCTGTGTGCTCTGATCGAGACAGAGGCTCATG 537
 Db 707 AGAGTGCCCGGAGCTGCGCATCAGCGCTATCATGCTGCTCTGATTGAGACCGCTTCATG 766
 Qy 538 ACCTCCTTC 546
 Db 767 ACTGCTTCC 775

RESULT 11
 BM683067/c
 LOCUS
 DEFINITION UI-E-E01-ajd-g-02-0-UI.s1 UI-E-E01 Homo sapiens cDNA clone
 UI-E-E01-ajd-g-02-0-UI 3', mRNA sequence.

ACCESSION BM683067
 VERSION BM683067.1 GI:18992963
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS 1 (bases 1 to 683)
 TITLE Ronaldo, M.F., Lennon, G. and Soares, M.B.

JOURNAL Normalization and subtraction: two approaches to facilitate gene
 MEDLINE discovery
 PUBMED 97044477

COMMENT
 Contact: Soares, MB
 8889548
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 1-44, >POLY A#simple_repeat (matched complement)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source
 1. .683
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-E01-ajd-g-02-0-UI"
 /tissue_type="fetal eye"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-E01"
 /notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-E01 is a normalized cDNA library containing the
 following tissue(s): fetal eye. The library was
 constructed according to Ronaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally
 into pT73-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 CGGTATACC. This library was created for the program, Gene
 Discovery in the Visual system, supported by National Eye
 Institute (NEI).
 TAG_TISSUE=human fetal eye
 TAG_LIB=UI-E-E01
 TAG_SEQ=CGGTATACC

ORIGIN

Query Match 67.98; Score 377; DB 12; Length 683;
 Best Local Similarity 85.5%; Pred. No. 3.2e-44;
 Matches 419; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
 Qy 57 GCCGGCGGATCGAGGATCCGGGAGCGGACTTCCAGTGTCTCCAGAGCGCGCGC 116
 Db 682 CTCAGGGGATCGGGGATCCGGGGCGGACTTCCAGTGTCTCCAGAGCGCGGNC 623
 Qy 117 CGCGGGCTGGCGGACCTTCCGGGCTTCTGTCGTCGCGCTGCGAGGACCTCTACAG 176
 Db 622 CGTGGGCTGGGGGACCTTCCGGGCTTCTGTCGTCGCGCTGCGAGGACCTCTACAG 563
 Qy 177 CATCTGCGCGCGCGACCGCACCGGGGTGCGGCTGCTCAACCTCAGGGACAGGTCT 236
 Db 562 CATCTGCGCGCGTGGCGACCGCGAGCGTGGCCATCGTCAACCTCAAGGACAGGTCT 503
 Qy 237 CTTCCCGAGTGGAGGCTTATTTCTGGGCTCCGAGCGGAGCTGAGCCGGGCGCGC 296
 Db 502 GTTCCCGAGTGGAGGCTCTGTTCTCAGGCTGTGAGGTCTGAGGCTGAGGCGGCGC 443
 Qy 297 CATCTTCTTTTCAGCGGAGAGATCTCTGAGCACCCCGCTGCGCCCGGAGAGGT 356
 Db 442 CATCTTCTTTTCAGCGGAGAGATCTCTGAGCACCCCGCTGCGCCCGGAGAGGT 383
 Qy 357 GTGCGAGGCTCCGACCCAGCGGCGCGCTGACCGACACTCTCGAGAGCTGCGC 416
 Db 392 GTGCGAGGCTCCGACCCAGCGGCGCGCTGACCGACACTCTCGAGAGCTGCGC 323
 Qy 417 GACGAGGCGCGCGCGCACCGGCGAGCGCTGTGCTGTGCGGCGAGCTGCTGGA 476
 Db 322 GACGAGGCTCCCTCGCGCACCGGCGAGCGCTCTCGTGTGCGGCGAGCTCTGGG 263
 Qy 477 GCAGGAGCGCGGAGTCCGCGCACCGCTCTGCTGTGCTCTCATCGAGAACAGCGTCT 536
 Db 262 GCAGGAGCGCGGAGTCCGCGCACCGCTCTGCTGTGCTCTCATCGAGAACAGCGTCT 203
 Qy 537 GACCTCCTTC 546
 Db 202 GACTGCTCC 193

RESULT 12

AW089583/c
 LOCUS
 DEFINITION AW089583 692 bp mRNA linear EST 15-OCT-1999
 similar to SW:CAIH HUMAN P39060 COLLAGEN ALPHA 1(XVII) CHAIN
 [CONTAINS: ENDOSTATIN], mRNA sequence.

ACCESSION AW089583

VERSION AW089583.1 GI:6046927

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 682)

AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

[REDACTED]

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:29:41 ; Search time 45.7778 Seconds
(without alignment)
1135.676 Million cell updates/sec

Title: US-09-938-391-4
Perfect score: 966
Sequence: 1 HTHQDFQLVHLVALNSPOP.....CRHAFVVLICIENSVMTSPSK 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1990s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	100.0	184	5	AA017430 Canine en
2	966	100.0	230	5	AA017429 Canine pr
3	959	99.3	184	3	AAY70265 Canine an
4	850	88.0	184	5	ABG31794 Murine en
5	835	86.4	181	4	AAU00898 Human end
6	835	86.4	182	3	AAY59622 Human end
7	835	86.4	182	3	AAY94323 Human end
8	835	86.4	182	3	AAB28399 Human end
9	835	86.4	182	4	AAU00897 Human end
10	835	86.4	182	5	AAU77951 Amino aci
11	835	86.4	183	2	AAY02113 SEQ ID 76
12	835	86.4	183	2	AAY08693 Human end
13	835	86.4	183	3	AAY70252 Human ang
14	835	86.4	183	3	AAY90771 Human ang
15	835	86.4	183	3	AAB16451 Human ang
16	835	86.4	183	3	AAB30493 Amino aci
17	835	86.4	183	4	AAB49379 Human end
18	835	86.4	183	4	AAU00896 Human end
19	835	86.4	183	5	ABB79901 Human end
20	835	86.4	183	5	AAM49503 Human end
21	835	86.4	183	5	AAM48895 Human end
22	835	86.4	183	5	AAU97132 Human end
23	835	86.4	183	6	AAU97553 Human end
24	835	86.4	195	3	AAW90874 Human HMW
25	835	86.4	216	3	AAB30495 Amino aci

26	835	86.4	275	5	AAU76689 Synthetic
27	835	86.4	310	5	AAU76688 Human col
28	835	86.4	513	5	ABG73586 Human End
29	835	86.4	682	5	ABP41878 Human ova
30	835	86.4	684	2	AAW26327 Human alp
31	835	86.4	684	2	AAY25113 Human alp
32	835	86.4	684	5	AAO17357 Human col
33	835	86.4	1301	2	AAW92296 Human alp
34	835	86.4	1336	2	AAY08594 Human col
35	835	86.4	1336	6	ABP96308 Human end
36	835	86.4	1516	5	ABP8617 Human col
37	835	86.4	1516	5	ABP8617 Human pan
38	834	86.3	180	4	AAU00899 Human HMW
39	834	86.3	193	3	AAW90877 Human end
40	831	86.0	183	4	AB49810 Human end
41	831	86.0	184	4	AB49380 Murine en
42	831	86.0	191	3	AAB28398 Murine en
43	831	86.0	191	5	AAU77950 Amino aci
44	831	86.0	207	5	ABE79902 Mouse end
45	828	85.7	184	2	AAW08689 Murine en

ALIGNMENTS

RESULT 1
AA017430
ID AA017430 standard; protein; 184 AA.

AC AA017430;

DT 19-JUL-2002 (first entry)

XX Canine endostatin.

XX Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
XX Poriasis; rheumatoid arthritis; retinopathy; macular degeneration;
XX corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
XX rubeosis; Osier-Webber Syndrome; myocardial angiogenesis;
XX plaque neovascularisation; telangiectasia; haemophilic joints;
XX angiofibroma; wound granulation; coronary collateral;
XX cerebral collateral; arteriovenous malformation;
XX ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
XX cyrostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
XX Gynaecological.

XX Canis familiaris.

XX EPI191036-A2.

PD 27-MAR-2002.

XX 24-AUG-2001; 2001EP-00307224.

XX 25-AUG-2000; 2000US-0227924P.

XX (PFIZ) PFIZER PROD INC.

XX Sheppard MG, Tong X;

XX WP1; 2002-354068/39.

XX N-ESDB; AAL46063.

PT An isolated nucleic acid molecule for the treatment of angiogenesis-
PT related disorder, such as cancers or diabetic retinopathy, encodes an
PT endostatin protein.

XX Claim 14; Fig 5; 56pp; English.

XX The present invention provides the protein and coding sequences of canine
CC pro-endostatin and endostatin. The sequences can be used in the treatment
CC and diagnosis of angiogenesis related disorders, including cancer,
CC rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,

CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC wound granulation, coronary collaterals, cerebral collaterals,
CC arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC neovascularisation, and fractures. The present sequence is the canine pro
CC -endostatin protein sequence
XX
SQ Sequence 184 AA;

Query Match 100.0%; Score 966; DB 5; Length 184;
Best Local Similarity 100.0%; Pred. No. 5.6e-107;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQLVHLVALNSPQPGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 1 HTHQDFQLVHLVALNSPQPGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
QY 61 VRRADRTGVPVNLDRDLVLPSPNEALFSGSEGOLKPGARIFSDGSDVLOHPAMPKRSVW 120
DB 61 VRRADRTGVPVNLDRDLVLPSPNEALFSGSEGOLKPGARIFSDGSDVLOHPAMPKRSVW 120
QY 121 HGSDFSGRRLLTDSYCETWRTTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 180
DB 121 HGSDFSGRRLLTDSYCETWRTTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 180
QY 181 SFSK 184
DB 181 SFSK 184

RESULT 2
AAO17429
ID AAO17429 standard; protein; 230 AA.

XX AAO17429;

XX 19-JUL-2002 (first entry)

XX Canine pro-endostatin.

XX Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
XX psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
XX corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
XX rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
XX plaque neovascularisation; telangiectasia; haemophiliac joints;
XX angiofibroma; wound granulation; coronary collateral;
XX cerebral collateral; arteriovenous malformation;
XX ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
XX cystic; antihypertensive; antidiabetic; ophthalmological;
XX synaecological.

XX Canis familiaris.

XX EP1191036-A2.

XX 27-MAR-2002.

XX 24-AUG-2001; 2001EP-00307224.

XX 25-AUG-2000; 2000US-0227924P.

XX (PTIZ) PFIZER PROD INC.

XX Sheppard MG, Tong X;

XX WPI; 2002-354069/39.

XX N-PSDB; AAL46062.

XX An isolated nucleic acid molecule for the treatment of angiogenesis-
XX related disorder, such as cancers or diabetic retinopathy, encodes an
XX endostatin protein.

PS Claim 14; Fig 3; 56pp; English.

XX The present invention provides the protein and coding sequences of canine
CC pro-endostatin and endostatin. The sequences can be used in the treatment
CC and diagnosis of angiogenesis related disorders, including cancer,
CC rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC wound granulation, coronary collaterals, cerebral collaterals,
CC arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC neovascularisation, and fractures. The present sequence is the canine pro
CC -endostatin protein sequence
XX

SQ Sequence 230 AA;

Query Match 100.0%; Score 966; DB 5; Length 230;
Best Local Similarity 100.0%; Pred. No. 7.8e-107;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQLVHLVALNSPQPGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 47 HTHQDFQLVHLVALNSPQPGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 106
QY 61 VRRADRTGVPVNLDRDLVLPSPNEALFSGSEGOLKPGARIFSDGSDVLOHPAMPKRSVW 120
DB 107 VRRADRTGVPVNLDRDLVLPSPNEALFSGSEGOLKPGARIFSDGSDVLOHPAMPKRSVW 166
QY 121 HGSDFSGRRLLTDSYCETWRTTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 180
DB 167 HGSDFSGRRLLTDSYCETWRTTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLICIENSVM 226
QY 181 SFSK 184
DB 227 SFSK 230

RESULT 3

AAAY70265

ID AAAY70265 standard; protein; 184 AA.

XX AAAY70265;

XX 06-JUN-2000 (first entry)

XX Canine angiogenesis inhibitor, endostatin.

XX Canine; immunoglobulin Fc fragment; endostatin; immunofusin;
XX angiogenesis; inhibitor; cytostatic; antirheumatoid; antihypertensive;
XX antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
XX vasotropic; vulnery; treatment; antiarteriosclerosis; tumour;
XX metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
XX ocular angiogenic disease; diabetic retinopathy; macular degeneration;
XX myocardial angiogenesis; plaque neovascularisation; telangiectasia;
XX wound granulation; keloid scar; gene therapy.

XX Canis familiaris.

XX WO200011033-A2.

XX 02-MAR-2000.

XX 25-AUG-1999; 99WO-USO19329.

XX 25-AUG-1998; 98US-0097883P.

XX (LEXI-) LEXINGEN PHARM CORP.

XX Lo X, Li Y, Gillies SD;

XX WPI; 2000-237616/20.

XX N-PSDB; AAZ51309.

PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin
 PT FC region, useful for treating conditions mediated by angiogenesis, such
 PT as rheumatoid arthritis, tumors and macular degeneration.

XX Example 8; Page 59-60; 68pp; English.

XX The patent discloses a DNA molecule encoding a fusion protein comprising
 CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis
 CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment
 CC having angiostatin activity, a collagen XVIII fragment having endostatin
 CC activity, or combinations of them. The fusion protein (immunofusin) is
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated
 CC by angiogenesis. Conditions that may be treated include solid tumors,
 CC blood born tumors, tumor metastasis, benign tumors including
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, rubeosis and Osler-Weber syndrome; myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
 CC endothelial cells, intestinal cells, atherosclerosis, scleroderma and
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
 CC gene therapy. The present sequence is a canine endostatin used in the
 CC construction of immunofusin containing canine immunoglobulin Fc fragment

XX Sequence 184 AA;

Query Match 99.3%; Score 959; DB 3; Length 184;
 Best Local Similarity 99.5%; Pred.No. 3.9e-106;
 Matches 183; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HTHQDFQLVHLVALNSPQPGMGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60
 Db 1 HTHQDFQLVHLVALNSPQPGMGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60
 QY 61 VRRADRTGVPVNNLRDEVLFPSSWEALFSGSEGQKPGARIFSPDGRDVLQHPAWPKSVW 120
 Db 61 VRRADRTGVPVNNLRDEVLFPSSWEALFSGSEGQKPGARIFSPDGRDVLQHPAWPKSVW 120
 QY 121 HGSDFSGRLTDSYCEWTWTEAPATQQAASLLAGRLLEQEAASCRHAFVLCIENSVM 180
 Db 121 HGSDFSGRLTDSYCEWTWTEAPATQQAASLLAGRLLEQEAASCRHAFVLCIENSVM 180

QY 181 SFSK 184
 Db 181 SFSK 184

RESULT 4
 ABG31794
 ID ABG31794 standard; protein; 184 AA.

XX AC ABG31794;

XX 05-DEC-2002 (first entry)

DE Murine endostatin polypeptide.

XX Mouse; endostatin; tumour; cancer; metastasis; cytostatic;
 KW antiangiogenic.

XX OS Mus sp.

XX PN WO200268457-A2.

XX PD 06-SEP-2002.

XX PF 27-FEB-2002; 2002WO-IT000119.

XX PR 27-FEB-2001; 2001IT-MI000394.

XX XX (UTMI-) UNIV MILANO.

XX
 PI
 XX
 DR
 XX

Chillemi F, Vicentinie LMT, Francescato P;

WPI; 2002-698655/75.

PT New peptide useful for the preparation of medicaments with antiangiogenic
 PT activity that may be used in treating tumors or metastases, comprises a
 PT sequence corresponding to fragments of human endostatin.

XX Disclosure; Fig 1; 24pp; English.

XX The invention relates to peptide comprising 20-50 amino acids with
 CC sequences corresponding to the human endostatin polypeptide sequence, its
 CC salt or non-toxic derivative. The peptides are useful in the preparation
 CC of medicaments with antiangiogenic activity which may be useful in
 CC treating tumors or metastases. This sequence represents a murine
 CC endostatin polypeptide

XX Sequence 184 AA;

Query Match 88.0%; Score 850; DB 5; Length 184;
 Best Local Similarity 85.3%; Pred. No. 4.4e-93;
 Matches 157; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQLVHLVALNSPQPGMGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60

Db 1 HSHRDFQPVHLVALNSPLSGMGIRGADFCQQAARAVGLAGTFRFLSSRLQDLYSI 60

QY 61 VRRADRTGVPVNNLRDEVLFPSSWEALFSGSEGQKPGARIFSPDGRDVLQHPAWPKSVW 120

Db 61 VRRADRAAVPIVNLKDELFFSWEALFSGSEGPKPGARIFSPGKDVLRHPTWPKSVW 120

QY 121 HGSDFSGRLTDSYCEWTWTEAPATQQAASLLAGRLLEQEAASCRHAFVLCIENSVM 180

Db 121 HGSDFSGRLTDSYCEWTWTEAPATQQAASLLAGRLLEQEAASCRHAFVLCIENSVM 180

QY 181 SFSK 184

Db 181 AFSK 184

RESULT 5

AAU00898
 ID AAU00898 standard; protein; 181 AA.

XX AC AAU00898;

XX 04-JUL-2001 (first entry)

DE Human Endostatin(TM) C-terminus minus 2 protein.

XX Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
 KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
 KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
 KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 KW retinopathy of prematurity; macular corneal graft rejection;
 KW neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KW haemophilic joint; angiofibroma; wound granulation; variant;
 KW C-terminus minus 2 protein.

XX OS Homo sapiens.

XX PN WO200119989-A2.

XX PD 22-MAR-2001.

XX PF 14-SEP-2000; 2000WC-US025166.

XX PR 14-SEP-1999; 99US-0153698P.

XX XX (ENTR-) ENTREMED INC.

AAV94323
ID AAV94323 standard; protein; 182 AA.
XX
AC AAV94323;
XX
DT 11-AUG-2000 (first entry)
XX
DE Human endostatin protein.
XX
KW Human; endothelial cell proliferation inhibitor; collagen XVIII;
KW angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;
KW vasotropic; dermatological; ophthalmological; vulvar; anti-
KW arteriosclerotic; antidiabetic; haemostatic; contraceptive;
KW ocular angiogenic disease; atherosclerosis; scleroderma;
KW myocardial angiogenesis; telangiectasia; angiofibroma; wound granulation.
XX
OS Homo sapiens.
XX
DE WO2000026368-A2.
XX
XX 11-MAY-2000.
XX
XX 01-NOV-1999; 99WO-US025605.
XX
XX 30-OCT-1998; 98US-0106343P.
XX
XX 20-MAY-1999; 99US-00315689.
XX
XX (CHIL-) CHILDRENS MEDICAL CENT.
XX
XX O'reilly MS, Folkman MJ;
XX
XX WPI; 2000-365617/31.
XX
XX N-PSDB; AAA27004.
XX
XX Novel endostatin capable of inhibiting endothelial cell proliferation and
XX angiogenesis, useful for treating angiogenesis-dependent cancers and as
XX birth control agents.
XX
XX Claim 2; Page 38; 68pp; English.
XX
XX The present sequence is an endostatin protein which is the carboxy
XX terminal protein of human collagen XVIII. Recombinant mouse endostatin
XX (20 mg/kg) was administered subcutaneously to mice implanted with Lewis
XX lung carcinomas. There was tumour mass regression non-detectable levels
XX after 12 days of therapy due to the angiogenesis inhibitory activity of
XX endostatin. Thus the protein is useful for treatment of angiogenesis-
XX dependent cancers. The polynucleotide and polypeptide sequences of this
XX endostatin are useful for treating and diagnosis of tumours, ocular
XX angiogenic diseases, Osler-Webber syndrome, myocardial angiogenesis,
XX plaque neovascularisation, telangiectasia, haemophilic joints,
XX angiofibroma and wound granulation, for treatment of diseases related to
XX excessive or abnormal stimulation of endothelial cells e.g. intestinal
XX adhesions, atherosclerosis, scleroderma. The protein may also be useful
XX as a birth control agent by reducing or preventing uterine
XX vascularisation. The gene for endostatin may be isolated from cells or
XX tissue that express high levels of endostatin, eg. tumour cells, by
XX generating cDNA from mRNA using reverse transcriptase and then amplifying
XX the DNA sequence
XX
XX Sequence 182 AA;
Query Match 86.4%; Score 835; DB 3; Length 182;
Best Local Similarity 85.1%; Pred. No. 2.7e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCFQQAAGLAGTFRAPLSRLQDLYSI 60
DB 1 HSHRDFQVLHLVALNSPLSGGMRGIRGADFCFQQAAGLAGTFRAPLSRLQDLYSI 60
QY 61 VRRADRTGVPVNLRLDEVLPSPWEALFSGSEGKLPFGARIFSDGKDVLRHTWPKQSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSPWEALFSGSEGKLPFGARIFSDGKDVLRHTWPKQSVW 120
QY 121 HGSDDPSGRLLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSFMT 180
DB 121 HGSDDPNGRLLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSFMT 180
QY 181 S 181

121 HGSDDPSGRLLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSFMT 180
DB 121 HGSDDPNGRLLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSFMT 180
QY 181 S 181
DB 181 A 181
RESULT 8
AAB28399
ID AAB28399 standard; protein; 182 AA.
XX
AC AAB28399;
XX
DT 19-FEB-2001 (first entry)
XX
DE Human endostatin.
XX
XX Human; endostatin; cytostatic; antiproliferative;
KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
KW cancer; vascularised solid tumour.
XX
OS Homo sapiens.
XX
XX WO2000064946-A2.
XX
XX 02-NOV-2000.
XX
XX 28-APR-2000; 2000WO-US011367.
XX
XX 28-APR-1999; 99US-0131432P.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX Thorpe PE, Brekken RA;
XX
XX WPI; 2000-687317/67.
XX
XX Immunogenic composition for the treatment and diagnosis of cancer
XX comprises an anti-VEGF (vascular endothelial growth factor) antibody
XX binding the same epitope as the monoclonal antibody ATCC PTA 1595.
XX
XX Example 10; Page 291-292; 298pp; English.
XX
XX The present invention relates to anti-Vascular Endothelial Growth Factor
XX (VEGF) antibodies that bind to the same epitope as the monoclonal
XX antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
XX the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
XX receptor VEGFR1. The present sequence is human endostatin. Endostatin may
XX be conjugated onto the anti-VEGF antibodies of the present invention. The
XX anti-VEGF antibodies of the present invention are useful for the
XX treatment and diagnosis of cancer, especially vascularised solid tumours
XX
XX Sequence 182 AA;
Query Match 86.4%; Score 835; DB 3; Length 182;
Best Local Similarity 85.1%; Pred. No. 2.7e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCFQQAAGLAGTFRAPLSRLQDLYSI 60
DB 1 HSHRDFQVLHLVALNSPLSGGMRGIRGADFCFQQAAGLAGTFRAPLSRLQDLYSI 60
QY 61 VRRADRTGVPVNLRLDEVLPSPWEALFSGSEGKLPFGARIFSDGKDVLRHTWPKQSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSPWEALFSGSEGKLPFGARIFSDGKDVLRHTWPKQSVW 120
QY 121 HGSDDPSGRLLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSFMT 180
DB 121 HGSDDPNGRLLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLICIENSFMT 180
QY 181 S 181

DB 181 A 181

RESULT 9
AAU00897
ID AAU00897 standard; protein; 182 AA.
XX
XX
AC AAU00897;
XX
XX
DT 04-JUL-2001 (first entry)
XX
XX
DE Human Endostatin(TM) C-terminus minus 1 protein.
XX
XX
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophiliac joint; angiofibroma; wound granulation; variant;
KW C-terminus minus 1 protein.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200119989-A2.
XX
XX
PD 22-MAR-2001.
XX
XX
PF 14-SEP-2000; 2000WO-US025166.
XX
XX
PR 14-SEP-1999; 99US-0153698P.
XX
XX
PA (ENTR-) ENTREMED INC.
XX
XX
PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
XX
XX
DR WPI; 2001-244802/25.
DR N-PSDB; AAS00897.
XX
XX
PT Producing Endostatin protein for treating angiogenesis mediated diseases
PT such as solid tumors, comprises recombinantly producing the protein using
PT an expression system, and recovering and purifying the protein.
XX
XX
PS Claim 5; Page 30; 67pp; English.
XX
XX
CC The sequence represents Human Endostatin(TM) C-terminus minus 1 protein,
CC a natural variant lacking the C-terminal amino acid of Endostatin(TM)
CC recovered from fermentations of Fichia pastoris cultures harbouring a
CC expression plasmid containing the Endostatin(TM) DNA sequence given in
CC AAS00867. The new method of the invention is useful for producing,
CC recovering and purifying Endostatin(TM) from biological sources, such as
CC biological fluids, tissues, cells, culture media, and fermentation media.
CC Endostatin(TM) is useful for treating angiogenesis mediated diseases such
CC as solid tumours, blood borne tumours, leukaemias, tumour metastases,
CC benign tumours, e.g. haemangioma, acoustic neuromas, neurofibromas,
CC trachomas, and pyogenic granulomas, rheumatoid arthritis, psoriasis,
CC ocular angiogenic diseases, e.g., diabetic retinopathy, retinopathy of
CC prematurity, macular degeneration, corneal graft rejection, neovascular
CC glaucoma, colon cancer, retrolental fibroplasia, rubeosis, Osler-Weber
CC Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC telangiectasia, haemophiliac joints, angiofibroma, and wound granulation.
CC Endostatin(TM) is also useful for treating disease of excessive or
CC abnormal stimulation of endothelial cells such as intestinal adhesions,
CC atherosclerosis, scleroderma and hypertrophic scars. Higher yields of
CC more purified, and biologically active Endostatin(TM) are obtained by the
CC new method. Endostatin(TM) can be stored in buffers for extended periods
CC of time, and also subjected to lyophilisation, while preserving
CC biological activity. Centrifugation of broth from fermentation steps in
CC production is avoided, preventing unwanted potential cellular lysis and

CC contamination with additional proteins, pigments, enzymes and other
CC cellular chemicals and debris
XX
XX
SQ Sequence 182 AA;
Query Match 86.4%; Score 835; DB 4; Length 182;
Best Local Similarity 85.1%; Pred. No. 2.7e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 1 HTHODFQVLHLVALNSPOGGMRGIRGADFCFOQARAAGLAGTFRFLSRLQDLYSI 60
DB 1 HSHRDFQVLHLVALNSPLSGMRGIRGADFCFOQARAAGLAGTFRFLSRLQDLYSI 60
QY 61 VRRADRTGVVNVNLRDEVLPSPWEALFSGSEGOLKPGARIFSGDGRDVLQHPAWPKRSW 120
DB 61 VRRADRAAVPINVNLKDELLFPSPWEALFSGSEGFLKPGARIFSGDKVLRHPTWPKSW 120
QY 121 HGGDPSGRRLTDSYCYETWTEAPAAAGQASSLLAGRLLEOEAASCRHAPVVLCIENSVM 180
DB 121 HGGDPNGRRRLTESYCYETWTEAPSAATGQASSLLGRLLCQSAASHAYIVLCIENSFMT 180
QY 181 S 181
DB 181 A 181

RESULT 10
AAU77951
ID AAU77951 standard; protein; 182 AA.
XX
XX
AC AAU77951;
XX
XX
DT 02-JUL-2002 (first entry)
XX
XX
DE Amino acid sequence for human endostatin.
XX
XX
KW Human; immunoconjugate; anti-vascular endothelial growth factor antibody;
KW anti-VEGF antibody; monoclonal antibody 2C3 ATCC PTA 1595; VEGF receptor;
KW VEGFR2; KDR/Flk-1; VEGFR1; PIt-1; angiogenesis; macular degeneration;
KW ocular neovascular disease; cancer; vascularised solid tumour; AIDS;
KW metastatic tumour; endothelial cell proliferation; inflammatory disorder;
KW atherosclerosis; diabetic retinopathy; corneal graft rejection;
KW acquired immune deficiency syndrome; infection; restenosis; fungal ulcer;
KW sickle cell anaemia; endometriosis; endostatin.
XX
XX
OS Homo sapiens.
XX
XX
PN AU200179401-A.
XX
XX
PD 06-DEC-2001.
XX
XX
PF 12-OCT-2001; 2001AU-00079401.
XX
XX
PR 28-APR-2000; 2000AU-00048049.
XX
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
XX
PI Thorpe PE, Brekken RA;
XX
XX
DR WPI; 2002-281368/33.
XX
XX
PT Immunoconjugate compositions for treating cancer by inhibiting
PT angiogenesis and for delivering a diagnostic agent to tumor, comprises
PT anti-vascular endothelial growth factor antibody attached to a biological
PT agent.
XX
XX
PS Example 10; Page 12-13 (Sequence listing); 299pp; English.
XX
XX
CC The present invention relates to antibody-based compositions comprising
CC an immunoconjugate such as anti-vascular endothelial growth factor (VEGF)
CC antibody (Ab) (or its antigen-binding fragment), attached to a biological
CC agent, where the Ab binds to the same epitope as the monoclonal antibody
CC (Mab) 2C3 ATCC PTA 1595, and significantly inhibits VEGF binding to the

CC VEGF receptor VEGFR2 (KDR/Flk-1) without inhibiting VEGF binding to the
CC VEGF receptor VEGFR1 (Flt-1). The compositions of the invention are
CC useful in therapy, and diagnosis, for inhibiting angiogenesis in an
CC animal having ocular neovascular disease or macular degeneration, and for
CC delivering a biological agent to a vascularised tumour. The compositions
CC can also be used for treating cancer and subjects at risk of developing,
CC a vascularised solid tumour, a metastatic tumour or metastases from a
CC primary tumour. The composition is useful for specifically inhibiting
CC VEGF-induced endothelial cell proliferation, without significantly
CC inhibiting VEGF-induced macrophage, osteoclast or chondrocyte function.
CC The compositions can be used for treating various diseases such as
CC inflammatory disorders, atherosclerosis, diabetic retinopathy,
CC restenosis, acquired immune deficiency syndrome (AIDS), blood borne
CC tumours, corneal graft rejection, Crohn's disease, fungal ulcers,
CC infections, sickle cell anaemia, and endometriosis. The present sequence
CC represents human endostatin. Endostatin may be attached or functionally
CC associated with anti-VEGF antibodies
XX
SQ Sequence 182 AA;

Query Match 86.4%; Score 835; DB 5; Length 182;
Best Local Similarity 85.1%; Pred. No. 2.7e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFQCFQQAARAGLAGTFFRAFLSSRLQDLYSI 60
Db 1 HSHRDFQPVHLHLVALNSPLSGGMRGIRGADFQCFQQAARAVGLAGTFFRAFLSSRLQDLYSI 60
QY 61 VRRADTGVVNVNLRDEVLPFSGSEGLKPGARIFSGDRVLOHPAPRKSVW 120
Db 61 VRRADRAAIVNLKDELLFPFSGSEGLKPGARIFSGDKVLRHPTWPKSVW 120
QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCTIENSFMT 180
Db 121 HGSDFNGRRLTESYCYETWTEAPATGQASSLLGRLLGQSAASCHHAYIVLCTIENSFMT 180
QY 181 S 181
Db 181 A 181

RESULT 11
AAY02113
ID AAY02113 standard; protein; 183 AA.
AC AAY02113;
DT 16-JUL-1999 (first entry)
DE SEQ ID 76 of WO9916889.
XX Angiostatin; endostatin; interferon; thrombospondin;
XX interferon-inducible protein; platelet factor 4; anti-angiogenic;
XX anti-tumor; multifunctional protein; angiogenic-mediated disease; cancer;
XX diabetic retinopathy; macular degeneration; arthritis;
XX tumor cell production.
XX Homo sapiens.
XX OS
XX WO9916889-A1.
XX PN
XX 08-APR-1999.
XX PD
XX 30-SEP-1998; 98WO-US020464.
XX PF
XX 01-OCT-1997; 97US-0060609P.
XX FR
XX (SEAR) SEARLE & CO G D.
XX PA
XX Bolanowski MA, Caparon MH, Casperson GF, Gregory SA, Klein BK;
XX Pi McKearn JP;
XX WPI; 1999-255098/21.

XX New multifunctional proteins useful for treating angiogenic-mediated
PT diseases.
PS Disclosure; Page 106-107; 121pp; English.
XX
CC The specification describes multifunctional proteins which comprise
CC combinations of angiostatin, endostatin, interferon, thrombospondin,
CC interferon-inducible protein and platelet factor 4, and have anti-
CC angiogenic and/or anti-tumor activity. The multifunctional protein may
CC exhibit useful properties such as having similar or greater biological
CC activity when compared to a single factor or by having improved half-life
CC or decreased adverse side effects, or a combination of these properties.
CC The proteins can be used for treating an angiogenic-mediated disease,
CC e.g. cancer, diabetic retinopathy, macular degeneration, or arthritis.
CC They can also be used for inhibiting the production of tumor cells
CC (characteristic of lung, breast, ovarian, prostate, pancreatic, gastric,
CC colon, renal, bladder cancers; melanoma, hepatoma, sarcoma and lymphoma)
CC in a patient and for inhibiting tumor growth. The present sequence is
CC used in the course of the invention
XX
SQ Sequence 183 AA;

Query Match 86.4%; Score 835; DB 2; Length 183;
Best Local Similarity 85.1%; Pred. No. 2.8e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFQCFQQAARAGLAGTFFRAFLSSRLQDLYSI 60
Db 1 HSHRDFQPVHLHLVALNSPLSGGMRGIRGADFQCFQQAARAVGLAGTFFRAFLSSRLQDLYSI 60
QY 61 VRRADTGVVNVNLRDEVLPFSGSEGLKPGARIFSGDRVLOHPAPRKSVW 120
Db 61 VRRADRAAIVNLKDELLFPFSGSEGLKPGARIFSGDKVLRHPTWPKSVW 120
QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCTIENSFMT 180
Db 121 HGSDFNGRRLTESYCYETWTEAPATGQASSLLGRLLGQSAASCHHAYIVLCTIENSFMT 180
QY 181 S 181
Db 181 A 181

RESULT 12
AAY08693
ID AAY08693 standard; protein; 183 AA.
AC AAY08693;
XX
XX 10-AUG-1999 (first entry)
DT
XX Human endostatin protein fragment.
XX DE
XX Plasmidogen; human; angiostatin; endostatin; gene therapy; vector;
XX anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
XX tumour growth; solid tumour; diabetic retinopathy; retina.
XX OS
XX Homo sapiens.
XX WO9926480-A1.
XX PN
XX 03-JUN-1999.
XX PD
XX 20-NOV-1998; 98WO-US024950.
XX PF
XX 20-NOV-1997; 97US-00975424.
XX FR
XX (GENE-) GENETIX PHARM INC.
XX PA (WASI) MASSACHUSETTS INST TECHNOLOGY.
XX PI Le Boulch P, Pawliuk RJ, Bachelot T;
XX

DR WPI; 1999-357696/30.
 DR N-PSDB; AAX77719.
 XX Anti-angiogenic gene therapy vectors.
 PS Disclosure; Page 74-75; 83pp; English.
 XX This invention describes a novel viral gene therapy vector comprising a
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from
 CC human or murine angiostatin, human or murine endostatin and angiogenesis-
 CC inhibiting fusions and fragments, where the viral vector is sufficiently
 CC attenuated for use in human gene therapy. The products of the invention
 CC have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological
 CC activity. The vector is used in gene therapy for inhibiting tumour growth
 CC in humans harbouring a solid tumour. The vector expresses an anti-
 CC angiogenic polypeptide. An additional use comprises treatment of diabetic
 CC retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis
 CC in the vicinity of the retina. The vector is administered to cells ex
 CC vivo and then administered to the patient
 XX Sequence 183 AA;
 SQ
 Query Match 86.4%; Score 835; DB 2; Length 183;
 Best Local Similarity 85.1%; Pred. No. 2.8e-91;
 Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
 QY 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFCFQQAARAGLAGTFFAFLSSRLQDLYSI 60
 Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFCFQQAARAVGLAGTFFAFLSSRLQDLYSI 60
 QY 61 VRRADRTGVVNLRLDEVLPFSWEALFSGSEGQLKPGARIFSGDRVLOHPAMPKRSVW 120
 Db 61 VRRADRAVPVNLKDELLFPSWEALFSGSEGFLKPGARIFSGDKVLRHFTWPQKSVW 120
 QY 121 HGSDPSGRRRLTDSYCTWTTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
 Db 121 HGSDPNRRRLTESYCTWTTEAPATGQASSLLGRLLGQAASCHHAYIVLCIENSFMT 180
 QY 181 S 181
 Db 181 A 181
 RESULT 13
 AAY70252
 ID AAY70252 standard; protein; 183 AA.
 AC AAY70252;
 XX
 DT 06-JUN-2000 (first entry)
 XX Human angiogenesis inhibitor, endostatin.
 DE Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;
 KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
 KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
 KW vasotropic; vulnery; treatment; antiarteriosclerosis; tumour;
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KW wound granulation; keloid scar; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FN WO200011033-A2.
 XX
 PD 02-MAR-2000.
 XX
 PF 25-AUG-1999; 99WO-US019329.
 XX
 PR 25-AUG-1998; 98US-0097883P.
 XX
 PA (LEXI-) LEXINGEN PHARM CORP.

XX Lo K, Li Y, Gillies SD;
 PI WPI; 2000-237616/20.
 DR N-PSDB; AAZ51291.
 XX Novel fusion protein of angiostatin or endostatin and an immunoglobulin
 PT Fc region, useful for treating conditions mediated by angiogenesis, such
 PT as rheumatoid arthritis, tumors and macular degeneration.
 XX Example 1; Page 41-42; 68pp; English.
 PS
 XX The patent discloses a DNA molecule encoding a fusion protein comprising
 CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis
 CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment
 CC having angiostatin activity, a collagen XVIII fragment having endostatin
 CC activity, or combinations of them. The fusion protein (immunofusin) is
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated
 CC by angiogenesis. Conditions that may be treated include solid tumours,
 CC blood born tumours, tumour metastasis, benign tumours including
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, rubecosis and Osler-Webber syndrome; myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
 CC endothelial cells, intestinal cells, and atherosclerosis, sclerodermal and
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
 CC gene therapy. The present sequence is a human endostatin used in the
 CC construction of immunofusin containing human immunoglobulin gamma (IgG)
 CC Fc fragment
 XX Sequence 183 AA;
 SQ
 Query Match 86.4%; Score 835; DB 3; Length 183;
 Best Local Similarity 85.1%; Pred. No. 2.8e-91;
 Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
 QY 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFCFQQAARAGLAGTFFAFLSSRLQDLYSI 60
 Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFCFQQAARAVGLAGTFFAFLSSRLQDLYSI 60
 QY 61 VRRADRTGVVNLRLDEVLPFSWEALFSGSEGQLKPGARIFSGDRVLOHPAMPKRSVW 120
 Db 61 VRRADRAVPVNLKDELLFPSWEALFSGSEGFLKPGARIFSGDKVLRHFTWPQKSVW 120
 QY 121 HGSDPSGRRRLTDSYCTWTTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
 Db 121 HGSDPNRRRLTESYCTWTTEAPATGQASSLLGRLLGQAASCHHAYIVLCIENSFMT 180
 QY 181 S 181
 Db 181 A 181
 RESULT 14
 AAY90771
 ID AAY90771 standard; protein; 183 AA.
 AC AAY90771;
 XX
 DT 22-AUG-2000 (first entry)
 XX Human angiogenesis inhibiting factor 1 protein.
 DE Human; angiogenesis inhibiting factor 1; IAP-1; tumour; antibody;
 KW Human; angiogenesis inhibiting factor 1; IAP-1; tumour; antibody;
 KW abnormal vessel disease.
 XX
 OS Homo sapiens.
 XX
 FN CN1244536-A.
 XX

PD 16-FEB-2000.
XX
XX 10-AUG-1998; 98CN-00117150.
XX
XX 10-AUG-1998; 98CN-00117150.
PR
XX (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.
PA
XX Yang Z, Guo W;
PI
XX WPI; 2000-388168/34.
XX
XX N-PSDB; AAA29884.
DR
XX Angiogenesis inhibiting factor 1 and its derivative useful for treating
PT tumors.
XX
XX Claim 1; Fig 5; 41pp; Chinese.
XX
XX The present sequence represents an angiogenesis inhibiting factor (I),
CC designated IAF-1. The present invention also describes: (1) preparation
CC of (I) and its derivative; (2) an IAF binding acceptor and its
CC preparation; and (3) an IAF antibody. (I) is useful for preparing new
CC biological preparations for effectively treating various tumors and
CC abnormal-vessel diseases. The IAF antibody is preferably a polyclonal
CC antibody, mosaic antibody, single stranded antibody and human originated
CC antibody
XX
XX Sequence 183 AA;
SQ
Query Match 86.4%; Score 835; DB 3; Length 183;
Best Local Similarity 85.1%; Pred. No. 2.8e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
Qy 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFCFOQARAAGLAGTFRFLSRLQDLYSI 60
Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFCFOQARAAGLAGTFRFLSRLQDLYSI 60
Qy 61 VRADRTGVPVNLDEVLPSPWEALFSGSEGQKPGARIFSDGDRVLOHPWPKSVW 120
Db 61 VRADRAAPVIVNLKDELLFPWEALFSGSEGQKPGARIFSDGDRVLOHPWPKSVW 120
Qy 121 HGSDPGRRLTDSYCTETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 121 HGSDPNGRRLTESYCTETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Qy 181 S 181
Db 181 A 181
RESULT 15
ID AAB16451 standard; protein; 183 AA.
XX
XX AAB16451;
XX
XX 27-OCT-2000 (first entry)
XX
XX Human endostatin protein sequence.
DE
XX Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;
KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
KW psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;
KW cerebral collateral; arteriovenous malformation; rubeosis; cancer;
KW diabetic retinopathy; arthritis; wound healing; peptic ulcer;
KW Helicobacter related disease; fracture; cat scratch fever.
XX
XX Homo sapiens.
OS
XX WO200032631-A2.
FN
XX
XX 08-JUN-2000.
XX

PF 06-DEC-1999; 99WO-US028897.
XX
XX 04-DEC-1998; 98US-00206059.
XX
XX (ENTR-) ENTREMED INC.
PA
XX Macdonald NJ, Sim KL;
XX
XX WPI; 2000-412290/35.
DR
XX New angiogenesis-inhibiting protein receptors, useful in methods for
PT treating diseases and processes that are mediated by angiogenesis, such
PT as solid tumors, psoriasis, scleroderma and myocardial angiogenesis.
XX
XX Disclosure; Fig 3; 100pp; English.
XX
XX This invention relates to angiogenesis-inhibiting protein receptors, and
CC the DNA sequences encoding them. Angiogenesis is the generation of new
CC blood vessels into a tissue, and normally occurs in wound healing, foetal
CC and embryonal development and the formation of the corpus luteum,
CC endometrium and placenta. Angiostatin is a protein (see AAB16450 and
CC AAA68202) involved in angiogenesis, and has an amino acid sequence
CC similar to that of a plasminogen fragment (see murine plasminogen
CC AAB16490). Angiostatin has the ability to inhibit angiogenesis.
CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and
CC AAA68203). Sequences AAA68242 and AAB16522 represent coding and protein,
CC and some of the peptides of the invention share homology with regions of
CC laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the
CC angiogenesis-inhibiting protein receptor fragments of the invention. The
CC peptides bind either angiostatin or endostatin and can be used in methods
CC for treating diseases and processes that are mediated by angiogenesis,
CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,
CC Crohn's disease, cerebral collateral, arteriovenous malformations,
CC rubeosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
CC Helicobacter related diseases, fractures, placenta and cat scratch
CC fever. They are useful for the detection and prognosis of cancer. DNA
CC sequences A628204-A628241 encode the peptides of the invention
XX
XX Sequence 183 AA;
SQ
Query Match 86.4%; Score 835; DB 3; Length 183;
Best Local Similarity 85.1%; Pred. No. 2.8e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
Qy 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFCFOQARAAGLAGTFRFLSRLQDLYSI 60
Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFCFOQARAAGLAGTFRFLSRLQDLYSI 60
Qy 61 VRADRTGVPVNLDEVLPSPWEALFSGSEGQKPGARIFSDGDRVLOHPWPKSVW 120
Db 61 VRADRAAPVIVNLKDELLFPWEALFSGSEGQKPGARIFSDGDRVLOHPWPKSVW 120
Qy 121 HGSDPGRRLTDSYCTETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 121 HGSDPNGRRLTESYCTETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Qy 181 S 181
Db 181 A 181
Search completed: March 26, 2004, 13:35:58
Job time : 47.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:34:07 ; Search time 14.6667 Seconds
(without alignments)
647.671 Million cell updates/sec

Title: US-09-938-391-4

Perfect score: 966

Sequence: 1 HTHQDFQLVHLVALNSPQP.....CRHAFVVLCTIENSVMYFSK 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/aaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/aaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/aaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/aaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/aaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	835	86.4	182	4	US-09-561-500-14
2	835	86.4	182	4	US-09-561-108-14
3	835	86.4	182	4	US-09-315-689-3
4	835	86.4	182	4	US-09-561-526-14
5	835	86.4	182	4	US-09-561-499-14
6	835	86.4	182	4	US-09-998-831-14
7	835	86.4	183	3	US-09-206-059-2
8	831	86.0	191	4	US-09-561-500-13
9	831	86.0	191	4	US-09-561-108-13
10	831	86.0	191	4	US-09-561-526-13
11	831	86.0	191	4	US-09-561-499-13
12	831	86.0	191	4	US-09-998-831-13
13	828	85.7	195	1	US-08-159-784-2
14	817	84.6	178	4	US-09-315-689-5
15	790	81.8	185	3	US-08-985-526-36
16	513	53.1	191	1	US-08-159-784-3
17	195.5	20.2	124	4	US-09-231-077D-10
18	188	19.5	123	4	US-09-231-077D-11
19	148	15.3	35	3	US-09-046-985-2
20	148	15.3	35	3	US-09-474-743-2
21	101	10.5	22	3	US-09-046-985-7
22	101	10.5	22	3	US-09-474-743-7
23	97	10.0	16	3	US-09-385-442-32
24	87	9.0	512	4	US-09-252-991A-17586
25	86	8.9	20	2	US-08-740-168A-1
26	86	8.9	20	3	US-09-349-429-1
27	86	8.9	20	4	US-09-315-689-1

Sequence 1, Appli
Sequence 1, Appli
Sequence 17378, A
Sequence 10595, A
Sequence 23863, A
Sequence 32078, A
Sequence 32350, A
Sequence 27, Appli
Sequence 6, Appli
Sequence 25606, A
Sequence 30853, A
Sequence 177, App
Sequence 177, App
Sequence 17508, A
Sequence 26179, A
Sequence 26045, A
Sequence 31937, A

ALIGNMENTS

RESULT 1

US-09-561-500-14
; Sequence 14, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-561-500-14

Query Match 86.4%; Score 835; DB 4; Length 182;
Best Local Similarity 85.1%; Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQLVHLVALNSPQGMGIRGADFCFQQAAGLACTFRAFLSSRLQDLYSI 60
Db 1 HSHRFQFVHLVALNSPLSGMGIRGADFCFQQAAGLACTFRAFLSSRLQDLYSI 60
QY 61 VRRADRTGVVYVNLREVLFPSSWEALFSGSGQLKPGARIFSPGGRDVLQHPWPKSVW 120
Db 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSGQLKPGARIFSPGGRDVLQHPWPKSVW 120
QY 121 HGSDPSGRRLTDSYCTETWTEAPATGASILLAGRLLEQEAASCRHAFVVLCTIENSVMY 180
Db 121 HGSDPSGRRLTDSYCTETWTEAPATGASILLAGRLLEQEAASCRHAFVVLCTIENSVMY 180
QY 181 S 181
Db 181 A 181

RESULT 2

US-09-561-108-14
; Sequence 14, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:

```

/ APPLICANT: Philip E. Thorpe
/
/ APPLICANT: Rolf A. Brekken
/
/ TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
/
/ FILE REFERENCE: 4001.002584
/
/ CURRENT APPLICATION NUMBER: US/09/561,108
/
/ CURRENT FILING DATE: 2000-04-28
/
/ PRIOR APPLICATION NUMBER: 60/131,432
/
/ PRIOR FILING DATE: 1999-04-28
/
/ NUMBER OF SEQ ID NOS: 44
/
/ SOFTWARE: PatentIn ver. 2.0
/
/ SEQ ID NO 14
/
/ LENGTH: 182
/
/ TYPE: PRT
/
/ ORGANISM: Artificial Sequence
/
/ FEATURE:
/
/ OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
/
/ OTHER INFORMATION: PEPTIDE
/
/ US-09-561-108-14

```

Query Match	86.4%;	Score	835;	DB	4;	Length	182;
Best Local Similarity	85.1%;	Pred.	No. 78-96;				
Matches 154; Conservative	14;	Mismatches	13;	Indels	0;	Gaps	0;
QY	1	HTHDDFVLVHLNALNSPQGGMRGIRGADFCFCQQAARAAGLAGTFFRAFLSSRLQDLYSI	60				
Db	1	HSHRDFVLVHLNALNSPLSGGMRGIRGADFCFCQQAARAVGLAGTFFRAFLSSRLQDLYSI	60				
QY	61	VRRDRGTGVVVNLRDLVPFSNEALFSGSEQLKPGARI FSDGRDYLOHPAKERSKVW	120				
Db	61	VRRADRAAVPTWNLDLLFPFSEALFSGSEGPLKPGARIFSDGKOVLRHPTWPQSKVW	120				
QY	121	HGSPDPGRRLTDSYCETWRTEAAPATQGASSLIAGRILLEQEAASCRHAFFVLCIENSVM T	180				
Db	121	HGSDPNGERLTESYCETWRTEAPSATQGASSLIGRLLGSAASHAYIVLCIENSFMT	180				
QY	181	S	181				
b	181	A	181				

```

RESULT 3
US-09-315-689-3
; Sequence 3, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1993-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-3

```

	Query Match	86.4%; Score 835; DB 4; Length 182;
	Best Local Similarity	85.1%; Pred.No.7e-96;
	Matches 154; Conservative	14; Mismatches 13; Indels 0; Gaps 0;
QY	1 HTHODFQLVHLVALNSPQGMRGIRGADFQCFCQAARAGLAGTFRFLSRLQDLYSI	60
DZ	: : : :	:
DB	1 HSHRDFQPVLHIVALNSPLSGMEGRGADFCFCQAARAVGLAGTFRFLSRLQDLYSI	60
QY	61 VRRARDTGVPVVNLRDEVLPFSWEALFGSGEOLKPGARI FSDGEDVLCIHPAPRSXW	120
DZ	: : : :	:
DB	61 VRADRRAA VPVTNKDELLFPSEALFGSGEGPKPGARIFSDGDVMRIHTPTWPQXS W	120
QY	121 HGSDPSGSRLLTDSCYTWRTEAPAATCQASSLAGLLEQEAAASHAFVILVC:ENSVM T	180
DZ	: : : :	:

```

Db      121  HGDPNRRRLTESCYCTWTEAFSATGQASSLLGQSSAAASHXAYIVLCIENSWMT 180
QY      181  S 181
      :
Db      181  A 181

      RESULT 4
      US-09-561-526-14
      ; Sequence 14, Application US/09561526
      ; Patent No. 6416758
      ; GENERAL INFORMATION:
      ; APPLICANT: Philip E. Thorpe
      ; APPLICANT: Rolf A. Brekken
      ; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
      ; FILE REFERENCE: 4001.002586
      ; CURRENT APPLICATION NUMBER: US/09/561,526
      ; CURRENT FILING DATE: 2000-04-28
      ; PRIOR APPLICATION NUMBER: 60/131,432
      ; PRIOR FILING DATE: 1999-04-28
      ; NUMBER OF SEQ ID NOS: 44
      ; SOFTWARE: PatentIn Ver. 2.0
      ; SEQ ID NO 14
      ; LENGTH: 182
      ; TYPE: PRT
      ; ORGANISM: Artificial Sequence
      ; FEATURE:
      ; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
      ; OTHER INFORMATION: PEPTIDE
      US-09-561-526-14

```

Query Match	86.4%;	Score	835;	DB 4;	Length	182;	
Best Local Similarity	85.1%;	Pred. No.	78-96;	13;	Indels	0;	
Matches	154;	Conservative	14;	Mismatches	0;	Gaps	0;
QY	1	HTHDFDLHLHVALNSP	GGMRGIRGADFCFCQQAARAAGLACTTFAFVLSRLQDLYSI	60			
Db	1	HSRDFQVTLHLHVALNS	PLSGMKGIRGADFCFCQQAARAVGLAGTTFRAFLSSRLQDLYSI	60			
QY	61	VRARDRTGVPVNIURDEV	LPSSWEALFSSGEQLKPGARIISFQGRDVLQHPAMPKRSVM	120			
Db	61	VRADRAAVPTVNLKDELL	FPSWEALFSSGEPLKPGARIISFQGVKVLRHPTWPKQSVW	120			
QY	121	HGSPDPSGRRLTDSYCET	WTREAPAAATQASSLLAGRLLEQEAASCHAFVVLCTIENSVM	180			
Db	121	HGSPNGRRLTESYCET	WTREAPSAATQASSLLGGRLLGQAASCHAYIVLCIENSFMT	180			
QY	181	S	181				
Db	181	A	181				

```

, RESULT 5
, US-09-561-499-14
, Sequence 14, Application US/09561499
, Patent No. 6524583
, GENERAL INFORMATION:
, APPLICANT: Philip B. Thorpe
, APPLICANT: Rolf A. Brekken
, TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
, FILE REFERENCE: 4001.002582
, CURRENT APPLICATION NUMBER: US/09/561,499
, CURRENT FILING DATE: 2000-04-28
, PRIOR APPLICATION NUMBER: 60/131,432
, PRIOR FILING DATE: 1999-04-28
, NUMBER OF SEQ ID NOS: 44
, SOFTWARE: PatentIn Ver. 2.0
, SEQ ID NO 14
, LENGTH: 162
, TYPE: PRT
, ORGANISM: Artificial Sequence
, FEATURE:
, OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

```

OTHER INFORMATION: PEPTIDE
US-09-561-499-14

Query Match 86.4%; Score 835; DB 4; Length 182;
Best Local Similarity 85.1%; Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 1 HSHRDFQVLHLVALNSPLSGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
QY 61 VRRADRTGVPVNLRLDEVLPFSPWEALFSGSEGQKPGARIFSDGDRDVLQHPAPRKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGQKPGARIFSDGDKVLRHPTWPKSVW 120
QY 121 HGSDDPSGRRLTDSYCTETWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
DB 121 HGSDDPNRRLLTESYCTETWTEAPAATQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180
QY 181 S 181
DB 181 A 181

RESULT 6
US-09-998-831-14

Sequence 14 Application US/09998831
Patent No. 6678941
GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/998,831
PRIOR FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 09/561,108
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-998-831-14

Query Match 86.4%; Score 835; DB 4; Length 182;
Best Local Similarity 85.1%; Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 1 HSHRDFQVLHLVALNSPLSGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
QY 61 VRRADRTGVPVNLRLDEVLPFSPWEALFSGSEGQKPGARIFSDGDRDVLQHPAPRKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGQKPGARIFSDGDKVLRHPTWPKSVW 120
QY 121 HGSDDPSGRRLTDSYCTETWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
DB 121 HGSDDPNRRLLTESYCTETWTEAPAATQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180
QY 181 S 181
DB 181 A 181

RESULT 7
US-09-206-059-2

Sequence 2, Application US/09206059

Patent No. 6201104
GENERAL INFORMATION:
APPLICANT: Macdonald, Nicholas
TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
Methods of Use
FILE REFERENCE: 05213-0370
CURRENT APPLICATION NUMBER: US/09/206,059
CURRENT FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 183
TYPE: PRT
ORGANISM: Homo sapiens
US-09-206-059-2

Query Match 86.4%; Score 835; DB 3; Length 183;
Best Local Similarity 85.1%; Pred. No. 7.1e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 1 HSHRDFQVLHLVALNSPLSGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
QY 61 VRRADRTGVPVNLRLDEVLPFSPWEALFSGSEGQKPGARIFSDGDRDVLQHPAPRKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGQKPGARIFSDGDKVLRHPTWPKSVW 120
QY 121 HGSDDPSGRRLTDSYCTETWTEAPAATQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
DB 121 HGSDDPNRRLLTESYCTETWTEAPAATQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180
QY 181 S 181
DB 181 A 181

RESULT 8
US-09-561-500-13

Sequence 13, Application US/09561500
Patent No. 6342219
GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002500
CURRENT APPLICATION NUMBER: US/09/561,500
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 13
LENGTH: 191
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-13

Query Match 86.0%; Score 831; DB 4; Length 191;
Best Local Similarity 83.7%; Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 8 HTHQDFQVLHLVALNTPLSGGMRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 67
QY 61 VRRADRTGVPVNLRLDEVLPFSPWEALFSGSEGQKPGARIFSDGDRDVLQHPAPRKSVW 120
DB 68 VRRADRAAVPIVNLKDEVLSPSNDLSLFGSQGQKPGARIFSDGDRDVLQHPAPRKSVW 127

QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 128 HGSDFSGRLMESYCYETWTEATTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187
QY 181 SFSK 184
Db 188 SFSK 191

RESULT 9
US-09-561-108-13
; Sequence 13, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-13

Query Match 86.0%; Score 831; DB 4; Length 191;
Best Local Similarity 83.7%; Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQPGMGRGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQVLHLVALNTPLSGMGRGIRGADFCQQAARAVGLSGTFRFLSSRLQDLYSI 67
QY 61 VRRADRTGVFVNLDRDEVLPSPWEALFSGSEGQLPGARIFSFQGRDVLQHPAPWPKSVW 120
Db 68 VRRADRTGVFVNLKDEVLSPWDSLFSGSQGLQPGARIFSFQGRDVLQHPAPWPKSVW 127
QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 128 HGSDFSGRLMESYCYETWTEATTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187
QY 181 SFSK 184
Db 188 SFSK 191

RESULT 10
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-526-13

Query Match 86.0%; Score 831; DB 4; Length 191;
Best Local Similarity 83.7%; Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQPGMGRGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQVLHLVALNTPLSGMGRGIRGADFCQQAARAVGLSGTFRFLSSRLQDLYSI 67
QY 61 VRRADRTGVFVNLDRDEVLPSPWEALFSGSEGQLPGARIFSFQGRDVLQHPAPWPKSVW 120
Db 68 VRRADRTGVFVNLKDEVLSPWDSLFSGSQGLQPGARIFSFQGRDVLQHPAPWPKSVW 127
QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 128 HGSDFSGRLMESYCYETWTEATTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187
QY 181 SFSK 184
Db 188 SFSK 191

RESULT 11
US-09-561-499-13
; Sequence 13, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-499-13

Query Match 86.0%; Score 831; DB 4; Length 191;
Best Local Similarity 83.7%; Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQPGMGRGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQVLHLVALNTPLSGMGRGIRGADFCQQAARAVGLSGTFRFLSSRLQDLYSI 67
QY 61 VRRADRTGVFVNLDRDEVLPSPWEALFSGSEGQLPGARIFSFQGRDVLQHPAPWPKSVW 120
Db 68 VRRADRTGVFVNLKDEVLSPWDSLFSGSQGLQPGARIFSFQGRDVLQHPAPWPKSVW 127
QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 128 HGSDFSGRLMESYCYETWTEATTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187
QY 181 SFSK 184
Db 188 SFSK 191

RESULT 12
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-526-13

Query Match 86.0%; Score 831; DB 4; Length 191;
Best Local Similarity 83.7%; Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQPGMGRGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQVLHLVALNTPLSGMGRGIRGADFCQQAARAVGLSGTFRFLSSRLQDLYSI 67
QY 61 VRRADRTGVFVNLDRDEVLPSPWEALFSGSEGQLPGARIFSFQGRDVLQHPAPWPKSVW 120
Db 68 VRRADRTGVFVNLKDEVLSPWDSLFSGSQGLQPGARIFSFQGRDVLQHPAPWPKSVW 127
QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 128 HGSDFSGRLMESYCYETWTEATTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187
QY 181 SFSK 184
Db 188 SFSK 191

RESULT 11
US-09-561-499-13
; Sequence 13, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-499-13

Query Match 86.0%; Score 831; DB 4; Length 191;
Best Local Similarity 83.7%; Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQPGMGRGIRGADFCQQAARAGLAGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQVLHLVALNTPLSGMGRGIRGADFCQQAARAVGLSGTFRFLSSRLQDLYSI 67
QY 61 VRRADRTGVFVNLDRDEVLPSPWEALFSGSEGQLPGARIFSFQGRDVLQHPAPWPKSVW 120
Db 68 VRRADRTGVFVNLKDEVLSPWDSLFSGSQGLQPGARIFSFQGRDVLQHPAPWPKSVW 127
QY 121 HGSDFSGRLTDSYCYETWTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 128 HGSDFSGRLMESYCYETWTEATTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187
QY 181 SFSK 184
Db 188 SFSK 191

RESULT 12
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Brekken
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
SEQUENCE CHARACTERISTICS:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 86.0%; Score 831; DB 4; Length 191;
Best Local Similarity 83.7%; Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVPLHLVALNTPSGMGRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 8 HTHQDFQVPLHLVALNTPSGMGRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 67
QY 61 VRRADRTGVVNLRLDEVLPFSSGQQLKPGARIFSPGDRDVLQHPAPKSVW 120
DB 68 VRRADRTGVVNLRLDEVLPFSSGQQLKPGARIFSPGDRDVLQHPAPKSVW 127
QY 121 HGSDFSGRLTDSYCTETWRTTETGATGQASSLLGRLLQEAASCHAFVVLCIENSFMT 180
DB 128 HGSDFSGRLTDSYCTETWRTTETGATGQASSLLGRLLQEAASCHAFVVLCIENSFMT 187
QY 181 SFSK 184
DB 188 SFSK 191

RESULT 13
US-08-159-784-2
Sequence 2, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 85.7%; Score 828; DB 1; Length 195;
Best Local Similarity 83.2%; Pred. No. 5.8e-95;
Matches 153; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVPLHLVALNTPSGMGRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 12 HTHQDFQVPLHLVALNTPSGMGRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 71
QY 61 VRRADRTGVVNLRLDEVLPFSSGQQLKPGARIFSPGDRDVLQHPAPKSVW 120
DB 72 VRRADRTGVVNLRLDEVLPFSSGQQLKPGARIFSPGDRDVLQHPAPKSVW 131
QY 121 HGSDFSGRLTDSYCTETWRTTETGATGQASSLLGRLLQEAASCHAFVVLCIENSFMT 180
DB 132 HGSDFSGRLTDSYCTETWRTTETGATGQASSLLGRLLQEAASCHAFVVLCIENSFMT 191
QY 181 SFSK 184
DB 192 SFSK 195

RESULT 14
US-09-315-689-5
Sequence 5, Application US/09315689
Patent No. 6346510
GENERAL INFORMATION:
APPLICANT: Folkman, Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
FILE REFERENCE: 05213-0229
CURRENT APPLICATION NUMBER: US/09/315,689
CURRENT FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 178
TYPE: PRT
ORGANISM: Homo sapiens
US-09-315-689-5

Query Match 84.6%; Score 817; DB 4; Length 178;
Best Local Similarity 85.9%; Pred. No. 1.2e-93;
Matches 152; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY 5 DFQVPLHLVALNTPSGMGRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 64
DB 1 DFQVPLHLVALNTPSGMGRGIRGADFCFQQAAGLAGTFRFLSSRLQDLYSI 60
QY 65 DRTGVVNLRLDEVLPFSSGQQLKPGARIFSPGDRDVLQHPAPKSVW 124
DB 61 DRAAVPIVNLKDELIPFSSGQQLKPGARIFSPGDRDVLQHPAPKSVW 120
QY 125 PSGRRLTDSYCTETWRTTETGATGQASSLLGRLLQEAASCHAFVVLCIENSFMT 181
DB 121 PSGRRLTDSYCTETWRTTETGATGQASSLLGRLLQEAASCHAFVVLCIENSFMT 177

RESULT 15
US-08-985-526-36
Sequence 36, Application US/08985526
Patent No. 6080728
GENERAL INFORMATION:
APPLICANT: Mixson, James A

TITLE OF INVENTION: CARRIER/DNA COMPLEXES CONTAINING DNA
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly, Bove, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: McMorow Jr., Robert G
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-36

Query Match 81.8%; Score 790; DB 3; Length 185;
Best Local Similarity 81.1%; Pred. No. 2.9e-90;
Matches 150; Conservative 16; Mismatches 17; Indels 2; Gaps 2;
QY 1 HTHQDFQLVHLVALNSPQPGMGRGIRGADFCQQARAAAGLAGTFRFLSSRLQDLYSI 60
Db 2 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFNAR-VGLSGTFRFLSSRLQDLYSI 60
QY 61 VRRADRTGVPVY-NLRDEVLPFSWEALFSGEGQKPGARIFSPDGRDVLQHPAPRKSV 119
Db 61 VRRADRGSPVIVONLRDEVLPFSWDSLPFSGQGQPGARIFSPDGRDVLQHPAPQRSV 120
QY 120 WHGSDPSGRRLTDSYCEWTREAPATGCQASSLLAGRLLEQEAASCRHAFVILCIENSYM 179
Db 121 WHGSDPSGRRLMESYCEWTRETTGATCQASSLLSGRLLEQEAASCHDSYIVILCIENSYM 180
QY 180 TSFSK 184
Db 181 TSFSR 185

Search completed: March 26, 2004, 13:39:22
Job time : 14.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:33:07 ; Search time 12.4444 Seconds
(without alignments)
1422.260 Million cell updates/sec

Title: US-09-938-391-4

Perfect score: 966
Sequence: 1 HTHQDFQVFLHLVALNSQP.....CRHAFVVLCLNENVTSPSK 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*

1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	835	86.4	684	A53019	collagen alpha 1(X)
2	831	86.0	1774	B56101	collagen alpha 1(X)
3	828	85.7	1315	A56101	collagen alpha 1(X)
4	526	54.5	1388	A53317	collagen alpha 1(X)
5	366	37.9	650	T22002	hypothetical prote
6	87	9.0	494	A83204	alginatase biosynth
7	82	8.5	995	H59332	RhoGAP protein hom
8	81.5	8.4	190	T36550	probable transcrip
9	81.5	8.4	633	G90704	probable Vgr prote
10	81.5	8.4	633	B85555	probable Vgr prote
11	81	8.4	312	A55461	NAD(P)-arginine AD
12	79.5	8.2	346	G83127	conserved hypothet
13	79.5	8.2	713	D90658	VgrG protein (limpo
14	79.5	8.2	713	D85509	hypothetical prote
15	79	8.2	672	T36083	hypothetical prote
16	78.5	8.1	244	A45729	sulfolipid biosynt
17	77	8.0	314	C23054	succinoglycan bios
18	77	8.0	314	F98231	succinoglycan bios
19	77	8.0	419	D42725	nitrite hydratase
20	77	8.0	800	A04338	probable Rhs acces
21	76	7.9	404	G90781	hypothetical prote
22	76	7.9	404	D85642	hypothetical prote
23	76	7.9	636	T10569	probable serine/th
24	76	7.9	732	AFO439	probable Rhs acces
25	75.5	7.8	227	A5128	carbohydrate kinas
26	75.5	7.8	319	D96159	hypothetical prote
27	75.5	7.8	336	T17408	rRNA (adenine-N6-)
28	75.5	7.8	385	T51127	GGPP synthase (imp
29	75.5	7.8	513	A45333	exopolysphatase

30	75.5	7.8	513	2	D91049	exopolysphatase
31	75.5	7.8	513	2	H58893	exopolysphatase
32	74.5	7.7	702	2	D50886	ygrE protein (limpo
33	74.5	7.7	714	2	B57332	Rhs element associ
34	74	7.7	126	2	A59177	hypothetical prote
35	74	7.7	636	2	JW0047	class I cytochrome
36	73.5	7.6	437	2	H58857	probable deaminase
37	73.5	7.6	723	2	AG1662	beta-glucosidases
38	73.5	7.6	807	2	F64844	ycds protein precu
39	73.5	7.6	807	2	F50787	probable outer mem
40	73.5	7.6	807	2	F56647	probable outer mem
41	73	7.6	427	2	A53798	58k membrane-assoc
42	73	7.6	449	2	H83944	glutamine syntheta
43	73	7.6	996	2	G87687	hypothetical prote
44	72.5	7.5	510	2	G87369	tryptophan halogen
45	72	7.5	286	2	E97241	hypothetical prote

ALIGNMENTS

RESULT 1

A53019
collagen alpha 1(XVIII) chain - human (fragment)
N;Contains: endostatin
C;Species: Homo sapiens (man)
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 15-Sep-2003
C;Accession: A53019
R;On, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen, Genomics 19, 494-499, 1994
A;Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localiz
A;Reference number: A53019; MUID:94245237; PMID:8188291
A;Accession: A53019
A;Molecule type: mRNA
A;Residues: 1-684 <OH>
A;Cross-references: GB:L22548; NID:9348908; PIDN:AAA51864.1; PID:5562794
A;Note: the cited accession number, L22548, is not in Genbank release 103
A;Comment: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifted
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C;Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivas
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of w
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: GDB:COL18A1
A;Cross-references: GDB:138752; OMIM:120328
A;Map position: 21q22.3-21q22.3
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly
F;1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>
F;1-59/Domain: collagenous (fragment) #status predicted <CO4>
F;74-115/Domain: collagenous #status predicted <CO5>
F;129-201/Domain: collagenous #status predicted <CO6>
F;212-244/Domain: collagenous #status predicted <CO7>
F;257-278/Domain: collagenous #status predicted <CO8>
F;262-264/Region: cell attachment (R-G-D) motif
F;286-340/Domain: collagenous #status predicted <CO9>
F;354-371/Domain: collagenous #status predicted <CO10>
F;502-684/Product: endostatin #status predicted <EST>
F;509-684/Region: multiplexin collagen carboxyl-terminal homologous

Query Match	86.4%	Score 835	DB 2	Length 684
Best Local Similarity	85.1%	Pred. No. 4.9e-74		
Matches 154	Conservative 14	Mismatches 13	Indels 0	Gaps 0
QY	1	HTHQDFQVFLHLVALNSQPQGMGRGIRGADFCQFQQAAGLAGTFRAPFLSRLQDLYSI	60	
Db	502	HSRDFQVFLHLVALNSPLSGMGRGIRGADFCQFQQAAGLAGTFRAPFLSRLQDLYSI	561	
QY	61	VRRADRTGVPVNRDLRDLVLPSPWEALFSGSEGQKPGARIFSGDGRDLVQLQHPAPRKSVW	120	
Db	562	VRRADRAAVFIVNLKDELLPSPWEALFSGSEGQKPGARIFSGDGRDLVQLQHPAPRKSVW	621	
QY	121	HGSDPSGRLLTDSYCEWTRTEAPATGQASSLLAGRLLEQAASCKHAPVLCIENSVMT	180	

Db 622 HGSDPGRRLTSCYCTWTEAPATGQASSLLGRLGQASCHAYIVLCIENSFMT 681
 Qy 181 S 181
 Db 682 A 682

RESULT 2
 collagen alpha 1(XVIII) chain precursor, long splice form - mouse
 N:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Oct-1995 #sequence revision 08-May-1998 #text change 15-Sep-2003
 C:Accession: B56101; C56101; S72450; S65595; P0675; A54072; A58816
 R:Rehm, M.; Pihlajaniemi, T.
 J. Biol. Chem. 270, 4705-4711, 1995
 A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue homologous to rat and Drosophila frizzled proteins.
 A:Reference number: A56101; MUID:95181468; PMID:7876242
 A:Accession: B56101
 A:Molecule type: mRNA
 A:Residues: 1-562 <REH1>
 A:Cross-references: GB:U11637; PIDN:AAC52179.1; PID:G618430
 A:Experimental source: splice form clone PE17.24
 A:Accession: C56101
 A:Molecule type: mRNA
 A:Residues: 1-239,487-562 <REH2>
 A:Cross-references: GB:U11637; PID:G618429
 A:Experimental source: splice form clones PB8.1, PE19, PE15.2
 R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
 submitted to the EMBL Data Library, August 1993
 A:Reference number: S72450
 A:Accession: S72450
 A:Molecule type: mRNA
 A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-1618
 A:Cross-references: EMBL:122545; NID:G348968; PIDN:AAA19787.1; PID:G511298
 R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
 A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
 A:Reference number: A58370, MUID:94240111; PMID:8183893
 A:Accession: S65595
 A:Molecule type: mRNA
 A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
 A:Cross-references: EMBL:122545
 R:Abe, N.; Muragaki, Y.; Yoshioaka, H.; Inoue, H.; Ninomiya, Y.
 Biochem. Biophys. Res. Commun. 196, 576-582, 1993
 A:Title: Identification of a novel collagen chain represented by extensive interruptions
 A:Reference number: P0675; MUID:94059075; PMID:8240330
 A:Accession: P0675
 A:Molecule type: mRNA
 A:Residues: 635-1774 <ABE>
 R:Rehm, M.; Hintikka, E.; Pihlajaniemi, T.
 J. Biol. Chem. 269, 13929-13935, 1994
 A:Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial structure
 A:Reference number: A54072; MUID:94245707; PMID:8188673
 A:Accession: A54072
 A:Molecule type: DNA; mRNA
 A:Residues: 1293-1403, 'R', 1405-1774 <REH3>
 A:Cross-references: GB:U03714; NID:G487733; PIDN:AAA20657.1; PID:G487734
 R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Billewicz, Z.; Roth, G.J.; Kleinman, H.K.; Kleinman, H.K.
 Cell 88, 277-285, 1997
 A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
 A:Reference number: A58816; MUID:97160848; PMID:9008168
 A:Accession: A58816
 A:Molecule type: protein
 A:Residues: 1591-1610 <ORE>
 A:Experimental source: hemangioendothelium cells
 A:Note: Inhibits endothelial cell proliferation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 A:Accession: A58371
 A:Molecule type: mRNA
 A:Residues: 1-928 <REH2>

C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u
 ay be useful in treating solid tumors.

C:Genetics:

A:Gene: MGI:Coll18a1

A:Cross-references: MGI:71175

A:Map position: 10:41.0

A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1595

A:Note: the list of introns is incomplete

C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly

F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status pre

F:1786-812/Domain: signal sequence #status predicted <SIG>

F:7823-896/Domain: collagenous #status predicted <COO1>

F:7921-1042/Domain: collagenous #status predicted <COO2>

F:1066-1148/Domain: collagenous #status predicted <COO3>

F:1163-1204/Domain: collagenous #status predicted <COO4>

F:11218-1290/Domain: collagenous #status predicted <COO5>

F:11301-1333/Domain: collagenous #status predicted <COO6>

F:1346-1363/Domain: collagenous #status predicted <COO7>

F:1351-1353/Region: cell attachment (R-G-D) motif

F:1377-1428/Domain: collagenous #status predicted <COO8>

F:1442-1459/Domain: collagenous #status predicted <COO9>

F:1591-1774/Product: endostatin #status predicted <EST>

F:1598-1774/Region: multiplexin collagen carboxyl-terminal homologous

F:354, 361, 947/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:699, 704, 1716/Binding site: carbohydrate (Ser) (covalent) #status predicted

F:910, 913, 1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 86.0%; Score 831; DB 2; Length 1774;

Best Local Similarity 83.7%; Pred. No. 3.8e-73;

Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

Qy	1	HTHODFQLVHLVALNSPQGGMRGIRGADFCFOQARAGLAGTFRFLSSRLQDLYSI	60
Db	1591	HTHODFQVPLVHLVALNTPLSGGMRGIRGADFCFOQARAVGLSGTFRFLSSRLQDLYSI	1650
Qy	61	VRADRTGVVNVLRDEVLPFSPWEALFSGSEQLKPGARIFSGDRDVLQHPANPRKSVW	120
Db	1651	VRADRGSPVPLNKLDEVLPSPWDSLFGSQQLPGARIFSGDRDVLQHPANPRKSVW	1710
Qy	121	HGSDPSGRRLTSCYCTWTEAPATGQASSLLGRLGQASCHAYIVLCIENSFMT	180
Db	1711	HGSDPSGRRLMESYCTWTEAPATGQASSLLGRLGQASCHAYIVLCIENSFMT	1770
Qy	181	SPSK 184	
Db	1771	SFSK 1774	

RESULT 3

A56101
 collagen alpha 1(XVIII) chain precursor, short splice form - mouse
 N:Contains: endostatin
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Oct-1995 #sequence revision 08-May-1998 #text change 15-Sep-2003
 C:Accession: A56101; A58371; S72450; S65595
 R:Rehm, M.; Pihlajaniemi, T.
 J. Biol. Chem. 270, 4705-4711, 1995
 A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue homologous to rat and Drosophila frizzled proteins.
 A:Reference number: A56101; MUID:95181468; PMID:7876242
 A:Accession: A56101
 A:Molecule type: mRNA
 A:Residues: 1-103 <REH1>
 A:Cross-references: GB:U11636; NID:G618427; PIDN:AAC52178.1; PID:G618428
 R:Rehm, M.; Pihlajaniemi, T.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
 A:Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous
 A:Reference number: A58371; MUID:94240112; PMID:8183894
 A:Accession: A58371
 A:Molecule type: mRNA
 A:Residues: 1-928 <REH2>

C;Date: 07-Jul-1995 #sequence revision 07-Jul-1995 #text_change 15-Sep-2003
C;Accession: A53317; A53146; S28778
R;Oh, S.P.; Kamegata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
Submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 28-697; 'L', 589-734, 'F', 736-751, 'R', 753-1315 <OH>
A;Cross-references: EMBL:L22545; NID:G348968; PIDN:AAA19787.1; PID:G511298
R;Oh, S.P.; Kamegata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 28-1315 <OH>
A;Cross-references: EMBL:L22545
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C;Comment: The different signal forms of collagen alpha 1(XVIII) may be involved in peri
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-235/Region: thrombospondin amino-terminal homologous
F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
F;327-353/Domain: collagenous #status predicted <CO1>
F;364-437/Domain: collagenous #status predicted <CO2>
F;462-583/Domain: collagenous #status predicted <CO3>
F;607-689/Domain: collagenous #status predicted <CO4>
F;704-745/Domain: collagenous #status predicted <CO5>
F;759-831/Domain: collagenous #status predicted <CO6>
F;842-874/Domain: collagenous #status predicted <CO7>
F;887-910/Domain: collagenous #status predicted <CO8>
F;892-984/Region: cell attachment (R-G-D) motif
F;918-969/Domain: collagenous #status predicted <CO9>
F;983-1000/Domain: collagenous #status predicted <CO10>
F;1132-1315/Product: endostatin #status predicted <EST>
F;1139-1315/Region: multiplexin collagen carboxyl-terminal homologous
F;126-488/Binding site: carbonyl (Asn) (covalent) #status predicted
F;172-228/Disulfide bonds: #status predicted
F;240-245, 2257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;451, 454, 594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 85.7%; Score 828; DB 2; Length 1315;
Best Local Similarity 83.2%; Pred. No. 5.2e-73;
Matches 153; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVFLHLVALNTPFGGVRGIRGADFCQFQARAGLAGTFRFLSSRLQDLYSI 60
DB 1132 HTHQDFQVFLHLVALNTPFGGVRGIRGADFCQFQARAGLAGTFRFLSSRLQDLYSI 1191

QY 61 VRRADRTGVPVNLRLDEVLPFSNEALFSGSEGQKPGARIFSPDGRDLVQLHPAPRWKSW 120
DB 1192 VRRADRTGVPVNLRLDEVLPFSNEALFSGSEGQKPGARIFSPDGRDLVQLHPAPRWKSW 1251

QY 121 HGSDPSGRLTDSYCEWTWRTEAPATQASLLAGRLLEQEAASCHAFVVLCIENSVM 180
DB 1252 HGSDPSGRLTDSYCEWTWRTEAPATQASLLAGRLLEQEAASCHAFVVLCIENSVM 1311

QY 181 SPSK 184
DB 1312 SPSK 1315

RESULT 4
A53317
Collagen alpha 1(XV) chain precursor - human
N;Alternate names: procollagen alpha 1(XV) chain
C;Species: Homo sapiens (man)

C;Date: 07-Jul-1995 #sequence revision 07-Jul-1995 #text_change 15-Sep-2003
C;Accession: A53317; A53146; S28778
R;Kivirikko, S.; Heinemann, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T
J. Biol. Chem. 269, 4773-4779, 1994
A;Title: Primary structure of the alpha chain of human type XV collagen and exon-intr
A;Reference number: A53317; MUID:94148920; PMID:8106446
A;Accession: A53317
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1388 <KIV>
A;Cross-references: GB:L25280
A;Note: nucleotide sequence and conceptual translation not complete
R;Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A;Title: The human alpha(XV) collagen chain contains a large amino-terminal non-triple
A;Reference number: A53146; MUID:94140817; PMID:8307960
A;Accession: A53146
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-9, 'S', 11-48, 'V', 50-94, 'A', 96-149, 'A', 151-203, 'V', 205-408, 'A', 410-569 <MUR
A;Cross-references: GB:D21230; NID:G415605; PIDN:BA04762.1; PID:G460703
R;Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), chara
A;Reference number: S28778; MUID:93066196; PMID:1279671
A;Accession: S28778
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 544-640, 'P', 642-811, 'P', 813-1252 <MYB>
C;Genetics:
A;Gene: GDB:COL15A1
A;Cross-references: GDB:132578; OMIM:120325
A;Map position: 9q21-9q22
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F;1216-1388/Region: multiplexin collagen carboxyl-terminal homologous

Query Match 54.5%; Score 526; DB 2; Length 1388;
Best Local Similarity 56.6%; Pred. No. 2.8e-43;
Matches 99; Conservative 27; Mismatches 45; Indels 4; Gaps 1;

QY 10 LHLVALNSPQGMGRGIRGADFCQFQARAGLAGTFRFLSSRLQDLYSI VRRADRTG 69
DB 1218 LHLVALNSPQGMGRGIRGADFCQFQARAGLAGTFRFLSSRLQDLYSI VRRADRTG 1273

QY 70 PVNLRDEVLPFSNEALFSGSEGQKPGARIFSPDGRDLVQLHPAPRWKSWHGSDPSGR 129
DB 1274 PVNLRDEVLPFSNEALFSGSEGQKPGARIFSPDGRDLVQLHPAPRWKSWHGSDPSGR 1333

QY 130 LTDSCYCEWTWRTEAPATQASLLAGRLLEQEAASCHAFVVLCIENSVMTSFSK 184
DB 1334 LTDSCYCEWTWRTEAPATQASLLAGRLLEQEAASCHAFVVLCIENSVMTSFSK 1388

RESULT 5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22002
R;White, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19500
A;Accession: T22002
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-650 <WIL>
A;Cross-references: EMBL:Z81079; PIDN:CAB03084.1; GSPDB:GN00019; CESP:F39H11.4
A;Experimental source: clone F39H11
C;Genetics:
A;Gene: CESP:F39H11.4
A;Map position: 1
A;Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

[illegible]

Query Match	9.0%;	Score 87;	DB 2;	Length 494;
Best Local Similarity	24.6%;	Prod. No. 1.2;		
Matches	45;	Conservative	20;	Mismatches 58; Indels 50; Gaps 7;

QY	4	QDPVLVHLVALNSPPQGMKGRGADF-----QCTQQAARAGLACTFPATUSSR	53
	:	:	:
	:	:	:
DB	39	KDFILLIGAVIWRYSNGGVHFLURGLFVHVYVYRRVRQLGSAADPSHFLMVTFSR	98
	:	:	:
	:	:	:
QY	54	LQDL-----YSIVRRADRTGVP-----VNLNREVLFPSS-WEAL-----	86
	:	:	:
	:	:	:
DB	99	IDALTAVTVRSVIREAIDSGYPTTVVCSIVMSDEVILVSLWEKKNPPDRVSLDFVRIP	158
	:	:	:
	:	:	:
QY	87	PSGSEGLKPGAIRFS-----PDGRDVT-QHPAWPFRKSVWVGSDPS-GRRLTDS	133
	:	:	:
	:	:	:
DB	159	GTGRDGLAYGFAIRSHRLFPDDDAVVAVIDGDTVLHDGVVYKTVFVFKLFENVGGLTNE	218
	:	:	:
	:	:	:
QY	134	YCE	136
	:		:
	:		:
DB	219	PCE	221

RESULT 7
H59432 RhoGAP protein homolog [imported] - human
C:Species: Homo sapiens (man)
C:Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 23-Sep-2002
C:Accession: H59432
R:Rhodes, S.
Submitted to GenBank, October 2001
A:Description: Novel human gene mapping to chromosome 13, similar to rat RhoGAP.
A:Reference number: H59432
A:Accession: H59432
A:Status: preliminary
A:Molecule type: DNA

```

A:Residues: 1-595 <RHO>
A:Cross-references: GS:NP_443083; PID:gl6445031; PIDN:NP_443083.1

Query Match      8.5%; Score 82; DB 2; Length 995;
Best Local Similarity 23.0%; Pred. No. 8.9;
Matches 37; Conservative 26; Mismatches 52; Indels 46; Gaps 7;

QY 52 SRLQDLISIVRRADRTGVP-----VVNLRDVLFPSEWALFSGEQQLXPG 97
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 36 SRVDLLTLPGRDRNGSPGTGMRNTTSESVLDTLSEPEVCSIHSSGGSRSQPG 95
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 98 -----AKIFSP-----GRDVLQHPAPRKSVWGHGDPGRRRTDSY--CETWR 139
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 96 QCCTDNPMVLDAPLVSSSLPQPPRDVLNHFHPK-----NEKPTRAKSFLKRMETLR 149
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 140 TEAPANTQOASS-----LLAGRLLECEAASCRHAFVVLGIE 175
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 150 GKGAGRHKGSGRTGGTVISGPMLOQEPESFK---AMQCIQ 187
   |||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 8
T36950
Probable transcription regulator - Streptomyces coelicolor
C.Species: Streptomyces coelicolor
C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C.Accession: T36950
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.J.
submitted to the EMBL Data Library, September 1999
A.Reference number: Z21607
A.Accession: T36950
A>Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A:Residues: 1-190 <SEE>
A:Cross-references: EMBL:AL109962; PIDN:CA853135.1; GSPDB:GN000070; SCOEDB:SCJ1.17
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ1.17

```

Query Match	8.4%	Score 81.5	DB 2	Length 190;
Best Local Similarity	25.1%	Pred. No. 1.4;		
Matches	46;	Conservative 18;	Mismatches 60;	Indels 59;
				Gaps 10;
Qy	11	HLVAINSFQCGHRTGADF-----	OCFOQARAAGLAGTFRAFLSSR	53
Db	18	HLAARATADGPREVEGKETMTSTTLPIAMPAQHRKLLRLAREVFPFGCARLFEEDR	77	
Qy	54	LQDLYSIVRRADRTGVPVNLN-----	DEVLPFSWE-----ALFSGSEGOLK	95
Db	78	PADRFTWV----RTGTVLVDIRVPDGRNAAVESLGHGELLGWSHFPFVYRHQGAEA--LS	132	
Qy	96	PGARIFSDGRDVL----QHPAMPKX-SVWHGSDPSGR-----	RLTDSYCEMTWTEAPA	144
Db	133	P-VRAWEFDKATVLTTCABEHFEFGRAIASWVGQVADRLQKTRIRLLDLY-----	APH	184
Qy	145	ATG	147	
Db	185	GSG	187	

RESULT 9
G90704
probable Vgr protein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050'
C.Species: Escherichia coli
C.Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C.Accession: G90704
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C (
sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA res. 8, 11-22, 2001
A.Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gei
A.Reference number: A99629; MUID:21156231; PMID:11259796
A.Accession: G90704
A.Status: preliminary
A.Molecule type: DNA

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and GenD
A:Reference number: A99629; UID:21156231; PMID:11258796
A:Accession: D90658
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA833659.1; PID:G13359692; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: EC80236

Query Match 8.2%; Score 79.5; DB 2; Length 713;
Best Local Similarity 24.4%; Pred. No. 10;
Matches 40; Conservative 18; Mismatches 59; Indels 47; Gaps 8;
QY 3 HQDFQLVHLVALNSPQGGNRGIRGADFCQQAAGAGTFRFLSRLQDLYSIVR 62
DB 247 HQDYQRTQYEV---YDYPGRFKGAHQNFARQMDGWRNNAEVARG--TSSRSPFIWP-GR 300
QY 63 RADRTGVPVNLRLDEVLFPSSWEALFS---GSEGQKPGARIFSDGDRDVLQHPAPWPKSV 119
DB 301 RIVLTGHPOANLNR-----NQVASELHGEQPOQAVPGR----- 335
QY 120 WHGSDPSGRRITDSYC-----ETWRT-----APAATGQASSLLAG 155
DB 336 -----GSGTTLNHNFAVIPADRTWRPQPLKPLVDGQSAVVTG 374

RESULT 14
DB5509
hypothetical protein 20267 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: D85509
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; UID:21074935; PMID:11206951
A:Accession: D85509
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <STO>
A:Cross-references: GB:AE005174; NID:G12512976; PIDN:AAG54536.1; GSPDB:GN00145; UWGP:202
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0267

Query Match 8.2%; Score 79.5; DB 2; Length 713;
Best Local Similarity 24.4%; Pred. No. 10;
Matches 40; Conservative 18; Mismatches 59; Indels 47; Gaps 8;
QY 3 HQDFQLVHLVALNSPQGGNRGIRGADFCQQAAGAGTFRFLSRLQDLYSIVR 62
DB 247 HQDYQRTQYEV---YDYPGRFKGAHQNFARQMDGWRNNAEVARG--TSSRSPFIWP-GR 300
QY 63 RADRTGVPVNLRLDEVLFPSSWEALFS---GSEGQKPGARIFSDGDRDVLQHPAPWPKSV 119
DB 301 RIVLTGHPOANLNR-----NQVASELHGEQPOQAVPGR----- 335
QY 120 WHGSDPSGRRITDSYC-----ETWRT-----APAATGQASSLLAG 155
DB 336 -----GSGTTLNHNFAVIPADRTWRPQPLKPLVDGQSAVVTG 374

RESULT 15
T36083
hypothetical protein SCE134.01c - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36083
R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999

A:Reference number: Z21596
A:Accession: T36083
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-672 <SAU>
A:Cross-references: EMBL:AL049661; PIDN:CA841199.1; GSPDB:GN00070; SCOEDB:SCE134.01c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE134.01c

Query Match 8.2%; Score 79; DB 2; Length 672;
Best Local Similarity 31.5%; Pred. No. 11;
Matches 39; Conservative 10; Mismatches 35; Indels 40; Gaps 8;
QY 48 AFLSSRLQDLYSIVRRADRTGVPVNLRLDEVLFPSSWEALFSGEGQKPGARIFSDGDRD 107
DB 465 AMFSEIRLG---VQVDRTLV-VTPLGRVVLDELEA-----EGPREWTY----- 504
QY 108 VLQHPAPWPKSVWHGSDPSGRRITDSYCETWRT-----APAATGQASSLLAGLLQEAASCRH 167
DB 505 -LLHSDWFTTQL-----DPDG-----WRLE-----SGPASARLT-RLLPDTATATRH 544
QY 168 AFVV 171
DB 545 RTVV 548

Search completed: March 26, 2004, 13:38:40
Job time : 14.4444 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:29:16 ; Search time 9.3333 Seconds
(without alignments)

1026.526 Million cell updates/sec

Title: US-09-938-391-4

Perfect score: 966

Sequence: 1 HTHQDFQLVHLVALNSQP.....CRHAFVVLCTIENSVMTSFSK 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	835	86.4	1516	1 CA1H HUMAN	P39060 homo sapien
2	831	86.0	1774	1 CA1H MOUSE	P39061 mus musculus
3	526	54.5	1388	1 CA1E HUMAN	P29059 homo sapien
4	87	9.0	494	1 ALG8 PSEAE	P24633 pseudomonas
5	85.5	8.9	556	1 GLI1 CHICK	P58788 gallus gall
6	82	8.5	995	1 NR13 HUMAN	Q9Y3M8 homo sapien
7	82	8.5	1233	1 NR13 HUMAN	Q14957 homo sapien
8	81	8.4	312	1 NR11 CHICK	P55806 gallus gall
9	77	8.0	419	1 P47K PSECL	P31521 pseudomonas
10	76.5	7.9	953	1 CAK4 MOUSE	Q8BHB0 mus musculus
11	76	7.9	7073	1 R1AB CVHSA	P59641 h replicase
12	75.5	7.8	512	1 PPX ECOLI	P29014 escherichia
13	73.5	7.6	326	1 PIM3 HUMAN	Q86V86 homo sapien
14	73.5	7.6	807	1 YCD5 ECOLI	P75907 escherichia
15	72.5	7.5	646	1 NA95 HUMAN	Q9ULX6 homo sapien
16	72	7.5	309	1 YHOC ECOLI	P45476 escherichia
17	72	7.5	5217	1 HT81 COCCA	Q01886 cochlilobol
18	71.5	7.4	356	1 MURG YERPE	Q8Z1E9 yersinia pe
19	71.5	7.4	1848	1 CCAA DROME	P91645 drosophila
20	71	7.3	999	1 MERK HUMAN	P12866 homo sapien
21	70.5	7.3	312	1 NR27 CHICK	P55807 gallus gall
22	70.5	7.3	317	1 GGH RAT	Q62867 rattus norv
23	70	7.2	185	1 NPW RAT	Q8K1M5 rattus norv
24	69.5	7.2	300	1 NARE CHICK	Q92080 gallus gall
25	69.5	7.2	415	1 VE2 PAPVE	P11329 european el
26	69.5	7.2	2799	1 EDD HUMAN	Q95071 homo sapien
27	69	7.1	676	1 UL06 HSV11	P10190 herpes simp
28	69	7.1	838	1 GLG8 SPRAW	Q821F0 streptomyce
29	68.5	7.1	319	1 HALL RAT	P15978 rattus norv
30	68.5	7.1	326	1 PIM3 MOUSE	P58750 mus musculus
31	68.5	7.1	326	1 PIM3 RAT	Q70444 rattus norv
32	68.5	7.1	428	1 NER3 BOVIN	Q97859 bos taurus
33	68.5	7.1	739	1 VNUC_EBOZ5	Q72142 ebola virus

ALIGNMENTS

RESULT 1

CALH HUMAN STANDARD; PRT; 1516 AA.

ID P39060: Q9UK38; Q9Y6Q7; Q9Y6Q8;
AC 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN COL18A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBF_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=98164096; PubMed=9503365;
RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Fihlajaniemi T.;
RT "Complete primary structure of two variant forms of human type XVIII
RT collagen and tissue-specific differences in the expression of the
RT corresponding transcripts";
RL Matrix Biol. 16:319-328(1998).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Tokoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lemann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]

RP SEQUENCE OF 834-1516 FROM N.A.
RX MEDLINE=94245237; PubMed=8189291;
RA Oh S.P., Warman M.L., Seidman M.F., Cheng S., Knoll J.H., Timmons S.,
RA Olsen B.R.;
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
RT and localization of the alpha 1(XVIII) collagen gene to mouse
RT chromosome 10 and human chromosome 21.";
RL Genomics 19:494-499(1994).
RN [4]

RP SEQUENCE OF 1334-1516 FROM N.A.
RX TISSUE=Placenta;
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT "Cloning and expression of human endostatin gene in Escherichia
RT coli.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [5]

RP INVOLVEMENT IN KNOBLOCH SYNDROME.

P18272 ebola virus
P17394 hepatitis b
Q92537 homo sapien
Q24246 drosophila
Q920P9 mesocricetu
P57842 pasteurella
Q87J64 vibrio para
Q89XK1 bradyrhizob
Q10958 mycobacteri
Q16842 homo sapien
P17564 mus musculu
P34544 caenorhabdi

```

RX MEDLINE=20400145; PubMed=10942434;
RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA Passos-Bueno M.R.;
RA "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT and tumor growth, plays a critical role in the maintenance of retinal
RT structure and in neural tube closure.";
RT Hum. Mol. Genet. 9:2051-2058(2000).
RN [6]
RN VARIANT ASN-1437.
RP MEDLINE=21518361; PubMed=1106364;
RX Iugnetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RX Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RX Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RA "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT for the development of prostatic adenocarcinoma.";
RT Cancer Res. 61:7375-7378(2001).
CC CC -!- FUNCTION: COL18A1 probably plays a major role in determining the
CC retinal structure as well as in the closure of the neural tube.
CC CC -!- FUNCTION: Endostatin potentially inhibits endothelial cell
CC proliferation and angiogenesis. May inhibit angiogenesis by
CC binding to the heparan sulphate proteoglycans involved in growth
CC factor signalling.
CC CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=NC-493;
CC IsoId=B39060-1; SequenceDisplayed;
CC Name=Short; Synonyms=NC1-303;
CC IsoId=B39060-2; Sequence=VSP_001155, VSP_001156;
CC CC -!- TISSUE SPECIFICITY: Present in multiple organs with highest levels
CC in liver, lung and kidney.
CC CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC CC -!- POLYMORPHISM: There is an association between a polymorphism in
CC position 1437 and prostate cancer. Heterozygous Asn-1437
CC individuals have a 2.5 times increased chance of developing
CC prostate cancer as compared with homozygous Asp-1437 individuals.
CC CC -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
CC [MIM:267750]; an autosomal recessive disorder defined by the
CC occurrence of high myopia, vitreoretinal degeneration with retinal
CC detachment, macular abnormalities and occipital encephalocele.
CC CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGEN WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; AF018081; AAC39658.1; -
DR EMBL; AF018082; AAC39659.1; -
DR EMBL; AL163302; CAB90482.1; -
DR EMBL; L22548; AAA51864.1; -
DR EMBL; AF184060; AAF01310.1; ALT_INT.
DR PDB; 1BNL; 02-DEC-98.
DR GlycoSuiteDB; P39060. -
DR Genew; HGNC:2195; COL18A1.
DR MIM; 120328; -
DR MIM; 267750; -
DR GO; GO:0005581; C:collagen; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR GO; GO:0007601; P:vision; TAS.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lect_gl.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 7.
DR Pfam; PF02210; TSPN;1.

```

	ProDom; P000007; Clg_helix; 1.	
DR	SMART; SMO0282; LamG; 1.	
DR	SMART; SMO0210; TSPN; 1.	
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;	
KW	Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;	
KW	Polyomorphisms; 3D-structure.	POTENTIAL.
FT	SIGNAL 1 23	
FT	CHAIN 24 1516	COLLAGEN ALPHA 1(XVIII) CHAIN.
FT	CHAIN 1334 1516	ENDOSTATIN.
FT	DOMAIN 221 409	TSP N-TERMINAL.
FT	DOMAIN 410 516	NONHELICAL REGION 1 (NC1).
FT	DOMAIN 517 550	TRIPLE-HELICAL REGION 1 (COL1).
FT	DOMAIN 551 560	NONHELICAL REGION 2 (NC2).
FT	DOMAIN 561 640	TRIPLE-HELICAL REGION 2 (COL2).
FT	DOMAIN 641 664	NONHELICAL REGION 3 (NC3).
FT	DOMAIN 665 786	TRIPLE-HELICAL REGION 3 (COL3).
FT	DOMAIN 787 809	NONHELICAL REGION 4 (NC4).
FT	DOMAIN 810 892	TRIPLE-HELICAL REGION 4 (COL4).
FT	DOMAIN 893 906	NONHELICAL REGION 5 (NC5).
FT	DOMAIN 907 948	TRIPLE-HELICAL REGION 5 (COL5).
FT	DOMAIN 949 961	NONHELICAL REGION 6 (NC6).
FT	DOMAIN 962 1034	TRIPLE-HELICAL REGION 6 (COL6).
FT	DOMAIN 1035 1044	NONHELICAL REGION 7 (NC7).
FT	DOMAIN 1045 1077	TRIPLE-HELICAL REGION 7 (COL7).
FT	DOMAIN 1078 1089	NONHELICAL REGION 8 (NC8).
FT	DOMAIN 1090 1111	TRIPLE-HELICAL REGION 8 (COL8).
FT	DOMAIN 1112 1118	NONHELICAL REGION 9 (NC9).
FT	DOMAIN 1119 1173	TRIPLE-HELICAL REGION 9 (COL9).
FT	DOMAIN 1174 1186	NONHELICAL REGION 10 (NC10).
FT	DOMAIN 1187 1204	TRIPLE-HELICAL REGION 10 (COL10).
FT	DOMAIN 1205 1516	NONHELICAL REGION 11 (NC11).
FT	CARBOHYD 68 68	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 129 129	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 164 164	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 591 591	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1329 1329	O-LINKED (GALNAc. .). /FTDISCAR_000150. BY SIMILARITY.
FT	DISULFID 1366 1506	BY SIMILARITY.
FT	DISULFID 1468 1498	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE 1095 1097	Missing (in isoform Short).
FT	VARSPLIC 1 180	/Ftid=VSP_001155. HTTEAGTLPAPTPSPSLGRPWAPLGPSPVPSS -> MA PCPCWPMPRRRLDLVLPLLVLLGVRAASASP (in isoform Short)
FT	VARSPLIC 181 215	/Ftid=VS_001156. D-> N (increased risk of developing prostate cancer). /Ftid=VAR_012709. F-> S (IN REF. 2). I-> V (IN REF. 2). V-> L (IN REF. 3). P-> R (IN REF. 3). P-> R (IN REF. 3). R-> L (IN REF. 3). P-> L (IN REF. 3). A-> P (IN REF. 3). L-> K (IN REF. 3). P-> A (IN REF. 3). P-> A (IN REF. 3). P-> PGPG (IN REF. 2). G-> GQ (IN REF. 3). R-> G (IN REF. 3). A-> G (IN REF. 3). LR -> CG (IN REF. 3). R-> T (IN REF. 4). S-> Y (IN REF. 4). SEQUENCE 1516 AA; 153840 MW; 3C70F29AA476EE76 CRC64; Query Match 86.4%; Score 835; DB 1; Length 1516; Best Local Similarity 85.1%; Pred.No.2.2e-72; Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps
FT	VARIANT 1437 1437	
FT	CONFLICT 428 428	
FT	CONFLICT 841 841	
FT	CONFLICT 877 877	
FT	CONFLICT 886 886	
FT	CONFLICT 912 912	
FT	CONFLICT 933 933	
FT	CONFLICT 975 975	
FT	CONFLICT 1064 1064	
FT	CONFLICT 1084 1084	
FT	CONFLICT 1120 1120	
FT	CONFLICT 1123 1123	
FT	CONFLICT 1126 1126	
FT	CONFLICT 1206 1206	
FT	CONFLICT 1304 1304	
FT	CONFLICT 1314 1314	
FT	CONFLICT 1323 1324	
FT	CONFLICT 1443 1443	
FT	CONFLICT 1483 1483	
FT	SEQUENCE 1516 AA; 153840 MW; 3C70F29AA476EE76 CRC64;	

Query Match 86.4%; Score 835; DB 1; Length 1516;
Best Local Similarity 85.1%; Pred. No. 2.2e-72;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQLVHLVALNSPQGMGIRGADFOCFQARAAAGLAGTFAFLSSRLQDLYSI 60
Db 1334 HSHRDQPVLHLVALNSPLSGWIRGADFOCQQAQAVGLAGTFAFLSSRLQDLYSI 1393
QY 61 VRADRTGTVNLRDEVLPFWNEALFSGSREGQLKPGARIPSPGRDVLQHPAPWPKSVW 120
Db 1394 VRADRAAIPVNLKDELLFPFWNEALFSGSEGLPKPGARIPSPGRDVLQHPAPWPKSVW 1453
QY 121 HGSDDPSGRLTDSYCTWRTAPATGQASSLLAGRLLEQAAACRAAFVVLCIENSVM 180
Db 1454 HGSDDPNGLRTSYCTWRTAPATGQASSLLAGRLLEQAAACRAAFVVLCIENSVM 1513
QY 181 S 181
Db 1514 A 1514

RESULT 2
ID CALH MOUSE
AC 239051; Q60672; Q61437; Q62001; Q62002; Q9JK63;
DT 01-FEB-1995 (Rel. 31, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].
GN COL18A1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M.V., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen
RT chain.";
RL J. Biol. Chem. 269:13929-13935 (1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3).
RX MEDLINE=96435922; PubMed=8938808;
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT "Characterization of the mouse gene for the alpha-1 chain of type
RT XVIII collagen (COL18A1) reveals that the three variant N-terminal
RT polypeptide forms are transcribed from two widely separated
RT promoters.";
RL Genomics 32:436-446 (1996).
RN [3]
RP SEQUENCE OF 1-1387 FROM N.A. (ISOFORM 3).
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M.V., Pihlajaniemi T.;
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238 (1994).
RN [4]
RP SEQUENCE OF 487-1774 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94240111; PubMed=8183893;
RA Oh S.P., Kamagaki Y., Muraagaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT "Isolation and sequencing of cDNAs for proteins with multiple domains
RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233 (1994).
RN [5]
RP SEQUENCE OF 1591-1774 FROM N.A.
RX MEDLINE=21217748; PubMed=11321448;
RA Jia S., Zhu F., Li H., He F., Xiu R.-J.;
RT "Anticancer treatment of endostatin gene therapy by targeting tumor
RT neovasculature in C57/BL mice.";
RL Clin. Hemorheol. Microcirc. 23:251-257 (2000).

[6]
RN CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE=97160848; PubMed=9008168;
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT growth.";
RL Cell 88:277-285 (1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX MEDLINE=98163982; PubMed=9501087;
RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT resolution.";
RL EMBO J. 17:1656-1664 (1998).
CC -!- FUNCTION: Endostatin potentially inhibits endothelial cell
CC proliferation and angiogenesis. May inhibit angiogenesis by
CC binding to the heparan sulfate proteoglycans involved in growth
CC factor signaling.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative promoter;
CC Comment=2 isoforms, 1 (shown here) and 3, are produced by use
CC of alternative promoters;
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1; Synonyms=NC1-764;
CC IsoId=P33061-3; Sequence=Displayed;
CC Name=2; Synonyms=Long, NC1-517;
CC IsoId=P33061-1; Sequence=VSP_008303;
CC Note=Produced by alternative splicing of isoform 1;
CC Name=3; Synonyms=Short, NC1-301;
CC IsoId=P33061-2; Sequence=VSP_001157, VSP_001158;
CC -!- PTM: Polines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (PACIT) FAMILY.
CC -!- SIMILARITY: Contains 1 frizzled (FZ) domain.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16898; AAA37434.1; -;
DR EMBL; U03714; AAA20657.1; -;
DR EMBL; U03715; AAC52901.1; -;
DR EMBL; U34608; AAC52901.1; JOINED.
DR EMBL; U34609; AAC52901.1; JOINED.
DR EMBL; U34610; AAC52901.1; JOINED.
DR EMBL; U34611; AAC52901.1; JOINED.
DR EMBL; U34612; AAC52901.1; JOINED.
DR EMBL; U34613; AAC52901.1; JOINED.
DR EMBL; U03716; AAC52901.1; JOINED.
DR EMBL; U03718; AAC52901.1; JOINED.
DR EMBL; U03715; AAC52902.1; -;
DR EMBL; U34607; AAC52902.1; JOINED.
DR EMBL; U34608; AAC52902.1; JOINED.
DR EMBL; U34609; AAC52902.1; JOINED.
DR EMBL; U34610; AAC52902.1; JOINED.
DR EMBL; U34611; AAC52902.1; JOINED.
DR EMBL; U34612; AAC52902.1; JOINED.
DR EMBL; U34613; AAC52902.1; JOINED.
DR EMBL; U34614; AAC52902.1; JOINED.
DR EMBL; U03716; AAC52902.1; JOINED.
DR EMBL; U03718; AAC52902.1; JOINED.
DR EMBL; U03715; AAC52903.1; -;
DR EMBL; U03716; AAC52903.1; JOINED.
DR EMBL; U03718; AAC52903.1; JOINED.
DR EMBL; U34607; AAC52903.1; JOINED.
DR EMBL; U34608; AAC52903.1; JOINED.

DR EMBL; U34609; AAC52903.1; JOINED.
 DR EMBL; U34610; AAC52903.1; JOINED.
 DR EMBL; U34611; AAC52903.1; JOINED.
 DR EMBL; U34612; AAC52903.1; JOINED.
 DR EMBL; U34613; AAC52903.1; JOINED.
 DR EMBL; U11636; AAC52178.1; --
 DR EMBL; U11637; AAC52179.1; --
 DR EMBL; U22545; AAC19787.1; --
 DR EMBL; AF257775; AAF69009.1; --
 DR PIR; A56101; A56101.
 DR PDB; 1KOE; 16-FEB-99.
 DR PDB; 1DYO; 11-APR-00.
 DR PDB; 1DY1; 21-JAN-01.
 DR MGD; MGI:88451; Coll18a1.
 DR GO; GO:0005604; C-basement membrane; IDA.
 DR GO; GO:0015255; P-angiotensin; IMP.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 8.
 DR Pfam; PF02210; TSPN; 1.
 DR ProDom; PRO00007; C1g_helix; 1.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS00308; FZ; 1.
 DR PROSITE; PS00308; FZ; 1.
 KW Extracellular matrix; Collagen; Glycoprotein; Signal; Hydroxylation;
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
 KW Alternative promoter usage; 3D-structure.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 1774 COLLAGEN ALPHA 1 (XVIII) CHAIN.
 FT CHAIN 1591 1774 ENDOSTATIN.
 FT DOMAIN 245 433 TSP N-TERMINAL.
 FT DOMAIN 365 482 FZ.
 FT DOMAIN 786 812 TRIPLE-HELICAL REGION 1 (NC1).
 FT DOMAIN 813 822 NON-HELICAL REGION 2 (NC2).
 FT DOMAIN 823 896 TRIPLE-HELICAL REGION 2 (COL2).
 FT DOMAIN 897 920 TRIPLE-HELICAL REGION 3 (NC3).
 FT DOMAIN 921 1042 TRIPLE-HELICAL REGION 3 (COL3).
 FT DOMAIN 1043 1065 NON-HELICAL REGION 4 (NC4).
 FT DOMAIN 1066 1148 TRIPLE-HELICAL REGION 4 (COL4).
 FT DOMAIN 1149 1162 NON-HELICAL REGION 5 (NC5).
 FT DOMAIN 1163 1204 TRIPLE-HELICAL REGION 5 (COL5).
 FT DOMAIN 1205 1217 NON-HELICAL REGION 6 (NC6).
 FT DOMAIN 1218 1290 TRIPLE-HELICAL REGION 6 (COL6).
 FT DOMAIN 1291 1300 NON-HELICAL REGION 7 (NC7).
 FT DOMAIN 1301 1333 TRIPLE-HELICAL REGION 7 (COL7).
 FT DOMAIN 1334 1345 NON-HELICAL REGION 8 (NC8).
 FT DOMAIN 1346 1369 TRIPLE-HELICAL REGION 8 (COL8).
 FT DOMAIN 1370 1376 NON-HELICAL REGION 9 (NC9).
 FT DOMAIN 1377 1428 TRIPLE-HELICAL REGION 9 (COL9).
 FT DOMAIN 1429 1441 NON-HELICAL REGION 10 (NC10).
 FT DOMAIN 1442 1459 TRIPLE-HELICAL REGION 10 (COL10).
 FT DOMAIN 1460 1774 NON-HELICAL REGION 11 (NC11).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 1623 1763 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 1725 1755 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1351 1353 Missing (in isoform 3).
 FT VARSPPLIC 1 459 Missing (in isoform 3).
 FT VARSPPLIC 460 486 AGRLLPVVCASPSDEGVCYFIPGAA -> MAPRWHLLDV
 FT VARSPPLIC 460 486 LSLVLLLVARSVAEP (in isoform 3).
 FT VARSPPLIC 240 486 Missing (in isoform 2).
 FT VARSPPLIC 1147 1147 P -> L (IN REF. 4).
 FT CONFLICT 1194 1194 P -> P (IN REF. 4).
 FT CONFLICT 1211 1211 A -> R (IN REF. 4).

Query Match 86.0%; Score 831; DB 1; Length 1774;
 Best Local Similarity 83.7%; Pred No. 6.4e-72;
 Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;
 QY 1 HTHQDFQLVHLVALNSPQGGMRGIRGADFCQFQARAAAGTAGTFRAPLSRLQDLYSI 60
 DB 1591 HTHQDFQFVLHLVALNTPSLSGMGRIRGADFCQFQARAVAGLSGTFRAFSLSRLODLYSI 1650
 QY 61 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGOLKPGARIFSGDRDVLQHPAPWPKSVW 120
 DB 1651 VRRADRGSPVIVNLKDEVLPFSWDSLFSGSQGLQPGARIFSGDRDVLRLHPAPWPKSVW 1710
 QY 121 HGSDFSGRLTDSYCTWRTEAPANTQOASSLLAGLEQEAASCEHAFVVLCTNSVMT 180
 DB 1711 HGSDFSGRLTDSYCTWRTEAPANTQOASSLLAGLEQEAASCEHAFVVLCTNSVMT 1770
 QY 181 SFSK 184
 DB 1771 SFSK 1774
 RESULT 3
 CALE HUMAN
 ID CALE HUMAN STANDARD; PET; 1388 AA.
 AC P39059;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(XV) chain precursor.
 GN COL15A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Umbilical cord;
 RX MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Heinamaki P., Rehn M.V., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 RT exon-intron organization in the 3' region of the corresponding
 RT Gene.";
 RL J. Biol. Chem. 269:4773-4779 (1994).
 RN [2]
 RP SEQUENCE OF 1-569 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=94140817; PubMed=8307960;
 RA Muragaki Y., Abe N., Nimmiya Y., Olsen B.R., Ooshima A.;
 RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
 RT non-triple helical domain with a tandem repeat structure and homology
 RT to alpha 1(XVIII) collagen.";
 RL J. Biol. Chem. 269:4042-4046 (1994).
 RN [3]
 RP SEQUENCE OF 544-1252 FROM N.A.
 RX MEDLINE=93066196; PubMed=1279671;
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
 RT "Identification of a previously unknown human collagen chain, alpha
 RT 1(XV), characterized by extensive interruptions in the triple-helical
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148 (1992).
 CC -1- TISSUE SPECIFICITY. Expressed predominantly in internal organs
 CC such as adrenal gland, pancreas and kidney.
 CC -1- PWM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

Db 39 KDFILLGAVIMRYSMGVHFLRGLMLFLHVYPPYRRVRVQLGSAADPSHVLMTVTR 98
QY 54 LQDL-----YSIVRADRTGVP-----VNLRLDEVLPSS-WEAL----- 86
Db 99 IDALTMTAVRSVIREAIDSGPTTVVCSIVMSDEVLVRSLEKWNPPDRVSLDFVTRIP 158
QY 87 FSGSEGOLKPGCARIFS-----FDGRDVLQHPAPWRKSVVHMGSDPPS-GRRLTDS 133
Db 159 GTGKEDGLAVGFRAISRHLRDDDDAVVAVIDGTVLDHGVVKKTPWFKLPNPGVGLTNE 218
QY 134 YCE 136
Db 219 FCE 221

RESULT 5

GLI_CHICK STANDARD; PRT; 556 AA.
AC P55878;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein GLI1 (GLI) (Fragment).
GLI1 OR GLI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97105842; PubMed=9948590;
RA Mariño V., Johnson R.L., Vortkamp A., Tabin C.J.;
RT "Sonic hedgehog differentially regulates expression of GLI and GLI3
during limb development."
RL Dev. Biol. 180:273-283(1996).
CC -!- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING
CC NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT
CC AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL
CC NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE
CC TRANSDUCTION OF SHH SIGNAL (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U60762; AAB51659.1; -.
CC HSSP; P08151; 2GLI.
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR SMART; SM00355; Znf C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS01557; ZINC_FINGER_C2H2_2; 5.
KW Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;
KW Nuclear protein; Repeat.
FT ZN_FING 247 272 C2H2-TYPE.
FT ZN_FING 280 307 C2H2-TYPE.
FT ZN_FING 313 337 C2H2-TYPE.
FT ZN_FING 343 368 C2H2-TYPE.
FT ZN_FING 374 399 C2H2-TYPE.
FT NON_TER 556 556
SQ SEQUENCE 556 AA; 60215 MW; 722D2AAS1CA4D98 CRC64;

Query Match 8.9%; Score 85.5; DB 1; Length 556;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 53; Conservative 20; Mismatches 75; Indels 53; Gaps 10;

QY 2 THODFOLVLELVALNSPQFGMRGIRGADFCQFOQARAAGLAGTTRAF-----LSRLQDL 57
Db 47 SHRGYGLV-----PGTEHPGG-----ADGSRFSFPRGAGKLGKRRALSISPLSDSSVDL 96
QY 58 YSIVRADRTGVPVNLRLDEVLPSS-WEAL-----FSGSEGOLKPGCARIFS----- 102
Db 97 QTIVRTSPNLVAFINRSRCASAGGSGYHLISLSTISPSLGYNPPQGGQGGQGFHTPTPL 156
QY 103 -----FDGRDVLQHPAPWRKSVVHMGSDPPSRR-----LTDSCYETWRTB-----APAA 145
Db 157 PPCSHETLSSRPGLLHPTPARGTIKHCQOLKERSLSPLTAKYPEE-KSEGDISSPAS 215
QY 146 TGOASSLLAGRL-----LQOE 161
Db 216 TGTQDFLL-GMLSVRDDLEKE 235

RESULT 6

SR13_HUMAN STANDARD; PRT; 995 AA.
AC Q9Y3M8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE STAR-related lipid transfer protein 13 (StARD13) (START domain-
DE containing protein 13) (46H23.2).
GN STARD13 OR GF650.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hunt A.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May function as a GTPase-activating protein.
CC -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC -!- SIMILARITY: Contains 1 START domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL049801; CAB42562.1; -.
CC EMBL; Z84483; CAC94774.1; -.
DR PIR; H59432; H59432.
DR Genew; HGNC:19164; STARD13.
DR InterPro; IPR008936; Rho_GAP.
DR InterPro; IPR000198; RHO_GAP.
DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhoGAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhoGAP; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS50236; RHO_GAP; 1.
DR PROSITE; PS50848; START; 1.
KW GTPase activation.
FT DOMAIN 545 750 RHO-GAP.
FT DOMAIN 781 989 START.
SQ SEQUENCE 995 AA; 111191 MW; 3F608FA94A4EF8BF CRC64;

Query Match 8.5%; Score 82; DB 1; Length 995;
Best Local Similarity 23.0%; Pred. No. 5.4;
Matches 37; Conservative 26; Mismatches 46; Gaps 7;

DR PRINTS, PRO0970; RIBTRNSFRASE.
 DR PROSITE, PS01291; ART; 1.
 KW Transferase; Glycosyltransferase; NAD; Signal; Zymogen.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 31
 FT CHAIN 32 266
 FT FT
 FT PROPEP 267 312
 FT ACT SITE 231 231 BY SIMILARITY.
 SQ SEQUENCE 312 AA; 35318 MW; B82980439BC904FC CRC64;
 Query Match 8.4%; Score 81; DB 1; Length 312;
 Best Local Similarity 26.3%; Pred. No. 1.8;
 Matches 50; Conservative 13; Mismatches 51; Indels 76; Gaps 12;
 QY 2 THQD-----FQLVLHLVALNSPOP-----GGWGRIGADFCFQQAAGLAG 44
 DB 128 SHQYIHSYHFKLHFLTQALFALRASOPRYVYVRGVRGIR-----FWTQR--GKSV 179
 QY 45 TFAFLSSRL-----QDLYSIYVRADRTGVFVNL-----RDEVLPFSWBAL----- 86
 DB 180 RFGQFTSLRKEATVNFQDTLVVKTG--YGVPIKQSPFPSEDEVLIPFVEFEVIN 237
 QY 87 FSGSEGG-----LPGAR.FSDGRDV-----LQHPAMP-RKSVW 120
 DB 238 FSDNRGSKVLQHSKGMKMTNCELLRPOGGOWGRGHQEVGLGSLPVLFCRRVW 297
 QY 121 HG-----SDP 125
 DB 298 EGLGHRGGDP 307
 RESULT 9
 P47K_PSECL STANDARD; PRT; 419 AA.
 AC P31521;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 47 kDa protein (P47K)
 OS Pseudomonas chlororaphis (Pseudomonas aureofaciens)
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=333;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B23;
 EX MEDLINE=91193202; PubMed=2013568;
 RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
 RA Beppu T.;
 RT "Cloning and characterization of genes responsible for metabolism of
 RT nitrite compounds from Pseudomonas chlororaphis B23."
 RL J. Bacteriol. 173:2465-2472(1991).
 CC -!- FUNCTION: THE PRESENCE OF P47K IS CRITICAL FOR THE EXPRESSION
 CC OF THE NITRILE HYDROLASE GENES. MAY STABILIZE OR ACTIVATE THE
 CC NITRILE HYDROLASE PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D90216; BAA14247.1; --
 DR F01; D42725; D42725.
 DR InterPro; IPR003495; COBw.
 DR Pfam; PF02492; COBw; 1.
 SQ SEQUENCE 419 AA; 46666 MW; FF5113800E27FF0C CRC64;
 Query Match 8.0%; Score 77; DB 1; Length 419;
 Best Local Similarity 24.0%; Pred. No. 6.1;

Matches 43; Conservative 26; Mismatches 76; Indels 34; Gaps 9;
 QY 26 IRGADFCQQAAGLAGTFRFLSSR-IQDLYSIYVRADRTGVFVNLDEVLPFSWE 84
 DB 141 VDSQFOALLSETDTVARADTEAHTSTRHLADI--LIEQVEYANVLNKRDLIDEPGYQ 198
 QY 85 ALFSGSGQLKPGARIFS-----PGRDVLQHPAWEKSVVHSGDPSGR 128
 DB 199 AVHAILAG-LNPSARINPMAHGNVALSSLDLTHFLDPLSLAASPGWKR--MEADITPAS 255
 QY 129 RLTDSCYET---WRTEPAATGQASSLLA-----GLLQEQA---ASCRHAFVLCIN 176
 DB 256 E-SDTYGTGVTWYRERAPFPQRLBFLQKPHNGRLLRSGYFWLASRHLIEGLAQS 313
 RESULT 10
 CAR4_MOUSE STANDARD; PRT; 953 AA.
 AC Q8BHB0; Q8BUT6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Caspase recruitment domain protein 4.
 GN CARD4.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Cerebellum, and Spleen;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito K., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Giamond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King S.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shinada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Atakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [2]
 RP SEQUENCE FROM N.A. AND VARIANT ALA-884.
 RC STRAIN=Czech II, and FVB/N; TISSUE=Breast cancer;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.R.,
 RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kettenan M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smallius D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Mazza M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Enhances caspase-9-mediated apoptosis. Induces NF-kappa-
CC B activity via RICK (CARDIAX, RIP2) and IKK-gamma. Confers
CC responsiveness to intracellular bacterial lipopolysaccharides
CC (LPS) (By similarity).
CC -!- SUBUNIT: Self-associates. Binds to caspase-9 and RICK by CARD-CARD
CC interaction (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Contains 1 CARD domain.
CC -!- SIMILARITY: Contains 1 NACHT domain.
CC -!- SIMILARITY: Contains 9 leucine-rich (LRR) repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AK082663; BAC38566.1; -
CC ENBL; AK089562; BAC40940.1; -
CC ENBL; BC042670; AHA42670.1; -
CC ENBL; BC043670; AHA43670.1; -
CC MGD; MG1:1341839; Card4.
CC InterPro: IPR001315; CARD.
CC InterPro: IPR007091; LRR_RNinh.
CC InterPro: IPR007111; NACHT_Ntfase.
CC Pfam; PF00619; CARD; 1.
CC Pfam; PF05729; NACHT; 1.
CC PROSITE; PS0209; CARD; 1.
CC PROSITE; PS50837; NACHT; 1.
CC Apoptosis; ATP-binding; Repeat; Leucine-rich repeat; Polymorphism.
FT DOMAIN 15 107
FT CARD.
FT NP_BIND 136 531
FT ATP (POTENTIAL).
FT REPEAT 17 42
FT LRR 1.
FT REPEAT 702 725
FT LRR 2.
FT REPEAT 727 750
FT LRR 3.
FT REPEAT 755 778
FT LRR 4.
FT REPEAT 783 806
FT LRR 5.
FT REPEAT 839 862
FT LRR 6.
FT REPEAT 867 890
FT LRR 7.
FT REPEAT 895 918
FT LRR 8.
FT REPEAT 923 946
FT LRR 9.
FT VARIANT 884 884
FT S -> A (in strain Czech II).
SQ SEQUENCE 953 AA; 107739 MW; 39C639621CEBIA58 CRC64;

Query Match 7.9%; Score 76.5; DB 1; Length 953;
Best Local Similarity 24.4%; Pred. No. 17;
Matches 47; Conservative 28; Mismatches 57; Indels 61; Gaps 11;
QY 2 THQDFQVHLVALNSPQPGNGRI-----RGADFQCFQQAARAGLAGTFRFLSRL 54
DB 158 TYMD--TLMLVGFNNELGSLGDLCLDHPSTGVNLNHEGTTFVFGDAGVGSMLLQRL 215
QY 55 QDL-----YSIVRRADRTGVPVNLRDVLPFSW-----EALF 87
DB 216 QSLWASGLRTSTAKFFHFRCEMSECKESD-----MLSLQD-LLFKHFCYEPQPEVFF 269
QY 88 SGRSGQLK-PGARIPSFQGRVLQHPAPKRSVHSGSDPSGRRLTOSYCTETWTEAPAT 146

DB 270 S---FLLRPHPTALFTFDGLDELH-----SDFDLRSRVPDSCC-FWEPAHPLVL 313
QY 147 GQASSLLAGRLLE 159
DB 314 --LANLLSRLX 324
RESULT 11
RIAB CVHSA STANDARD; PRT; 7073 AA.
ID RIAB CVHSA STANDARD; Q80BV7; Q80BV8; Q80E51;
AC PS9641; Q80C0; Q80BV7; Q80BV8; Q80E51;
DT 10-OCT-2003 (rel. 42, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last sequence update)
DT 15-MAR-2004 (rel. 43, Last annotation update)
DE Replicase polyprotein lab (pp1AB) [ORF1AB] [Includes: Replicase
DE polyprotein 1a (pp1a) (ORF1A)] [Contains: leader protein; p65 homolog;
DE NSP1 (EC 3.4.24.-) (Papain-like proteinase) (FL-PRO); 3C-like
DE proteinase (EC 3.4.24.-) (3CLP) (NSP2); NSP3; NSP4; NSP5;
DE NSP6; Growth factor-like peptide (GFL) (NSP7); RNA-directed RNA
DE polymerase (EC 2.7.7.48) (RdRp) (NSP9); Helicase (Hel) (NSP10);
DE NSP11; NSP12; NSP13.
OS Human coronavirus (strain SARS) (HCoV-SARS) (SARS-COV)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=227859;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Isolate Urbani;
RC MEDLINE=22660724; PubMed=12730500;
RA Rota P.A., Oberste M.S., Monroe S.S., Nix W.A., Campagnoli R.,
RA Icenogle J.P., Penaranda S., Bankamp B., Maher K., Chen M.-H.,
RA Tong S., Tamin A., Lowe L., Frace M., Derisi J.L., Chen Q., Wang D.,
RA Raman D.D., Peret T.C.T., Burns C., Ksiazek T.G., Rollin P.E.,
RA Sanchez A., Liffick S., Holloway B., Limor J., McCaustland K.,
RA Olsen-Rasmussen M., Fouchier R., Guenther S., Osterhaus A.D.M.E.,
RA Crosten C., Pallansch M.A., Anderson L.J., Bellini W.J.,
RT "Characterization of a novel coronavirus associated with severe acute
RT respiratory syndrome."
RL Science 300:1394-1399(2003).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Isolate Tor2;
RC MEDLINE=22660725; PubMed=12730501;
RA Marra M.A., Jones S.J.M., Astell C.R., Holt R.A., Brooks-Wilson A.,
RA Butterfield Y.S.N., Khatri J., Asano J.K., Barber S.A., Chan S.Y.,
RA Cloutier A., Coughlin S.M., Freeman D., Girn N., Griffith O.L.,
RA Leach S.R., Mayo M., McDonald H., Montgomery S.B., Pandoh P.K.,
RA Petrescu A.S., Robertson A.G., Schein J.E., Siddiqui A., Smallius D.E.,
RA Stott J.M., Yang G.S., Plummer F., Anderson A., Artsob H., Bastien N.,
RA Bernard K., Booth T.F., Bowness D., Czib M., Drebot M., Fernando L.,
RA Flick R., Garbutt M., Gray M., Grolla A., Jones S., Feldmann H.,
RA Meyers A., Kabani A., Li Y., Normand S., Stroher U., Tipples G.A.,
RA Tyler S., Vogrig R., Ward D., Watson B., Brunham R.C., Kraiden M.,
RA Patric M., Skowronski D.M., Upton C., Roper R.L.,
RT "The genome sequence of the SARS-associated coronavirus."
RL Science 300:1399-1404(2003).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=Isolate CUHK-Su10, and isolate CUHK-W1;
RC MEDLINE=22737955; PubMed=12853594;
RA Tsui S.K.W., Chim S.S.C., Lo Y.M.D.,
RT "Coronavirus genomic-sequence variations and the epidemiology of the
RT severe acute respiratory syndrome."
RL New Engl. J. Med. 349:187-188(2003).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN=Isolate G250, isolate sz3, and isolate sz16;
RC MEDLINE=22913660; PubMed=12958366;
RA Guan Y., Zheng B.J., He Y.O., Liu X.L., Zhuang Z.X., Cheung C.L.,
RA Luo S.W., Li P.H., Zhang L.J., Guan Y.J., Butt K.M., Wong K.L.,
RA Chan K.W., Lim W., Shortridge K.F., Yuen K.Y., Peiris J.S.M.,
POON L.L.M.;
RT "Isolation and characterization of viruses related to the SARS

RT coronavirus from animals in southern China.";
 RN Science 302:276-278(2003).
 [5]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate HKU-39849;
 RC MEDLINE=22758472; PubMed=12876307;
 RA Zeng F.Y., Chan C.W., Chan M.N., Chen J.D., Chow K.Y.C., Hon C.C.C.,
 RA Rui R.K.H., Li J., Li V.Y.Y., Wang C.Y., Wang P.Y., Guan Y., Zheng B.,
 RA Poon L.L.M., Chan K.H., Yuen K.Y., Peiris J.S.M., Leung F.C.;
 RT "The complete genome sequence of severe acute respiratory syndrome
 RT coronavirus strain HKU-39849 (HK-39).";
 RL Exp. Biol. Med. 228:866-873(2003).
 [6]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate Sin2500, Isolate Sin2677, Isolate Sin2679,
 RC Isolate Sin2774, and Isolate Sin2748;
 RX MEDLINE=223667074; PubMed=12781537;
 RA Ruan Y., Wei C.L., Ling A.E., Vega V.B., Thoreau H., Se Thoe S.Y.,
 RA Chia J.-M., Ng P., Chiu K.P., Lim L., Zhang T., Chan K.P., Oon L.E.L.,
 RA Ng M.L., Leo S.Y., Ng L.F.P., Ren E.C., Stanton L.W., Long P.M.,
 RA Liu E.T.;
 RT "Comparative full-length genome sequence analysis of 14 SARS
 RT coronavirus isolates and common mutations associated with putative
 RT origins of infection.";
 RL Lancet 361:1779-1785(2003).
 [7]
 RN ERRATUM.
 RP Ruan Y., Wei C.L., Ling A.E., Vega V.B., Thoreau H., Se Thoe S.Y.,
 RA Chia J.-M., Ng P., Chiu K.P., Lim L., Zhang T., Chan K.P., Oon L.E.L.,
 RA Ng M.L., Leo S.Y., Ng L.F.P., Ren E.C., Stanton L.W., Long P.M.,
 RA Liu E.T.;
 RL Lancet 361:1832-1832(2003).
 [8]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate BJ01, Isolate BJ02, Isolate BJ03, Isolate BJ04,
 RC and Isolate GD01;
 RA Qin E., Zhu Q., Yu M., Fan B., Chang G., Si B., Yang B., Peng W.,
 RA Liang T., Liu B., Deng Y., Liu H., Zhang Y., Wang C., Li Y., Gan Y.,
 RA Li X., Lu F., Tan G., Yang R., Cao W.S., Wang J., Chen W., Cong L.,
 RA Deng Y., Dong W., Han Y., Hu W., Lei M., Li C., Li G., Li H.,
 RA Li S., Li S., Li W., Lin W., Liu J., Liu Z., Lu H., Ni P.,
 RA Qi Q., Sun Y., Tang L., Tong Z., Wang J., Wang X., Wu Q., Xi Y.,
 RA Xu Z., Yang L., Ye C., Ye J., Zhang B., Zhang F., Zhang J., Zhang X.,
 RA Zhou J., Yang H.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 [9]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate TW1;
 RA Yeh S.-H., Kao C.-L., Tsai C.-Y., Liu C.-J., Chen D.-S., Chen P.-J.;
 RT "The complete genome of SARS coronavirus clone TW1.";
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 [10]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate FRA;
 RA Eickmann M., Becker S., Klenk H.-D., Doerr H.W., Stadler K.,
 RA Censini S., Guidotti S., Masignani V., Scarselli M., Mora M.,
 RA Donati C., Han J., Song H.C., Abrignani S., Covacci A., Rappuoli R.;
 RT "SARS virus is a close relative of type II coronaviruses.";
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 [11]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate Frankfurt 1;
 RC Thiel V., Hertzog T., Putics A., Ivanov K.A., Schelle B., Bayer S.,
 RA Schneider B., Weiland H., Weisbrich B., Ziebuhr J.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 [12]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate TWC;
 RA Yang J.-Y., Lin J.-H., Chiu S.-C., Wang S.-F., Lee S.-C., Lin Y.-C.,
 RA Hsu C.-K., Chen H.-Y., Chang J.-G., Chen P.-J., Su I.-J.;
 RT "Genomic sequence of SARS isolate from the first fatal case in
 RT Taiwan.";
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
 RP STRAIN=Isolate HSR 1;
 RC Canducci P., Clementi M., Poli G., Vicenzi E.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 [14]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate Taiwan TC1, Isolate Taiwan TC2, and Isolate Taiwan TC3;
 RC Chang J.-G.C., Lin T.-H., Chen C.-M., Lin C.-S., Chan W.-L.,
 RA Shin M.-C.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 [15]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate TWJ, Isolate TWK, Isolate TWS,
 RC and Isolate TWY;
 RA Shu H.-Y., Wu K.M., Tsai S.F.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 [16]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Isolate AS;
 RA Balotta C., Corvasce S., Violin M., Galli M., Moroni M.,
 RA Vigevani G.M., Ruan Y.J., Salemi M.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 [17]
 RN SEQUENCE OF 1-322 FROM N.A.
 RP STRAIN=Isolate ZJ-HZ01;
 RC Wang Z., Cheng S., Zhang Y.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 [18]
 RN SEQUENCE OF 1-507; 1655-5170 AND 6903-7073 FROM N.A.
 RP STRAIN=Isolate Shanghai 1Y;
 RA Yuan Z., Zhang X., Hu Y., Lan S., Wang H., Zhou Z., Wen Y.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 [19]
 RN SEQUENCE OF 497-1674 AND 5158-6911 FROM N.A.
 RP STRAIN=Isolate Shanghai QXC;
 RC Yuan Z., Zhang X., Hu Y., Lan S., Wang H., Zhou Z., Wen Y.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 [20]
 RN SEQUENCE OF 4993-5127 FROM N.A.
 RP STRAIN=Isolate Vietnam;
 RA Emery S., Erdman D.D., Peret T.C.T., Ksiazek T.G.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 [21]
 RN SEQUENCE OF 4993-5136 FROM N.A.
 RP STRAIN=Isolate Taiwan;
 RA Lin J.-H., Chiu S.-C., Yang J.-Y., Wang S.-F., Chen H.-Y.;
 RT "Detection of a novel human coronavirus in a severe acute respiratory
 RT syndrome patient in Taiwan.";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 [22]
 RN CHARACTERIZATION OF HELICASE
 RX MEDLINE=22890296; PubMed=12317423;
 RA Tanner J.A., Watt R.M., Chai Y.-B., Lu L.-Y., Lin M.C., Peiris J.S.,
 RA Poon L.L.M., Kung H.-F., Huang J.-D.;
 RT "The severe acute respiratory syndrome (SARS) coronavirus
 RT NPase/helicase belongs to a distinct class of 5' to 3' viral
 RT helicases.";
 RL J. Biol. Chem. 278:39578-39582(2003).
 [23]
 RN 3D-STRUCTURE MODELING OF 3241-3540, AND CHARACTERIZATION.
 RX MEDLINE=22690393; PubMed=12746549;
 RA Anand K., Ziebuhr J., Wadhwani P., Mesters J.R., Hilgenfeld R.;
 Query Match 7.9%; Score 76; DB 1; Length 7073;
 Best Local Similarity 22.1%; Pred. No. 2e+02;
 Matches 36; Conservative 24; Mismatches 73; Indels 30; Gaps 4;
 QY 1 HTTQDFQLVHLVALNS-----PQGGVGRGIRGADFCFOQARAAGLAGTFFRAFLSSRL 54
 Db 1613 HTLDESEFLGRYMSALNHTKKWFPQVGGTSTKWDNNCYLSSVLLALQQLVEKFNAPAL 1672
 QY 55 QDLYSIVRRADRTGVPEVNLNRDEVLPFSWEALFSGSEGQLKPGARIFPSFGDGRVLQHPAW 114

1673 QEAYYRAGDAANFCAILII-----AYSNKTVGSLG-----DYRETMTW--- 1715

115 PRKSWMGSDPSGRRLTDSYCETWRTEAPAAATGQASSLLAGRL 157

1712 ---LLQHANLESARKVNLNVCKGCGKTTLLTGVEAVWMTGL 1751

RESULT 12

PPX_ECOLI

ID PPX_ECOLI STANDARD; PRT; 512 AA.

AC P29014; P76981;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-APR-1993 (Rel. 45, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Exopolypolysphatase (EC 3.6.1.11) (Exopolypase) (Metaphosphatase).

GN PPX OR B2502 OR C3020 OR Z3765 OR ECS3364.

OS Escherichia coli,

OS Escherichia coli O6, and

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

NCBI_TaxID=562, 217992, 83334;

[1] SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.

RC STRAIN=K12;

RC MEDLINE=93107072; PubMed=83801070;

RA Akiyama M., Crooke E., Kornberg A.;

RT "An exopolypolysphatase of *Escherichia coli*. The enzyme and its ppX

RL gene in a polyphosphate operon.";

RL J. Biol. Chem. 268:633-639(1993).

[2]

RC SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RA MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of *Escherichia coli* K-12.";

RL Science 277:1453-1474(1997).

[3]

RC SEQUENCE FROM N.A.

RC STRAIN=K12;

RC MEDLINE=97349980; PubMed=9205837;

RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,

RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,

RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,

RA Oshima T., Oyama S., Saito N., Sanger J., Satoh Y., Sivasubram S.,

RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,

RA Yamagata S., Horiuchi T.;

RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*

RL - K12 genome corresponding to 50.0-68.8 min on the linkage map and

RL analysis of its sequence features.";

RL DNA Res. 4:91-113(1997).

[4]

RC SEQUENCE FROM N.A.

RC STRAIN=06:H1 / CFT073 / ATCC 700928;

RC MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,

RA Mobley H.L.T., Sonnenberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence

RL of uropathogenic *Escherichia coli*."

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

[5]

RC SEQUENCE FROM N.A.

RC STRAIN=0157:H7 / EDL933 / ATCC 700927;

RA MEDLINE=21074935; PubMed=11208551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

Db 292 SCGALRGV-LYMEGRVFRHQDVRSTASSLANQVHIDSEQARRVLDITMQMYSQWREQQ 350
 QY 143 P-ANTGASSLLAGRLLEQFAA-----SCRHAFVVLVCISNMTFSK 184
 Db 351 PKLAHPQLEALLRWAAHMEVLGNINHSGLHRHSAYILQNSDLPGFNQ 398

RESULT 13
 PIM3 HUMAN
 ID PIM3 HUMAN STANDARD; PRT; 326 AA.
 AC Q86V86;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Serine/threonine-protein kinase pim-3 (EC 2.7.1.37).
 GN PIM3.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.D., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Faney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra X.A.,
 RA "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP IDENTIFICATION FROM RSTS.
 RX MEDLINE=22682943; PubMed=12798037;
 RA Chichester C., Nikitin F., Ravarini J.-C., Lisacek F.;
 RT "Consistency checks for characterizing protein forms."
 RL Comput. Biol. Chem. 27:29-35 (2003).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC -1- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. PIM
 subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC ENBL: BC052239; AAH52239.1; ALT INIT.
 DR InterPro: IPR007119; Prot_kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR008271; Ser_thr_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TyKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 FT Phosphorylation.
 FT DOMAIN 40 293 PROTEIN_KINASE.
 FT NP_BIND 46 54 ATP (BY SIMILARITY).
 FT BINDING 69 69 ATP (BY SIMILARITY).
 FT ACT_SITE 170 170 BY SIMILARITY.
 SQ SEQUENCE 326 AA; 35863 MW; 41FDFDD2467A162 CRC64;
 Query Match 7.6%; Score 73.5; DB 1; Length 326;
 Best Local Similarity 26.1%; Pred. No. 9.9;
 Matches 42; Conservative 12; Mismatches 82; Indels 25; Gaps 5;
 QY 12 LVALNSPQPGMGIRGADPQCQQAARAGLACTPAFLSSRLQDLYSVRADTGVFV 71
 Db 120 LVLRDEP-----AQDLDFITRGALDSELRFFAQVL-----AAVRCHSCGVVH 168
 QY 72 VNLREVLFFPSWEA----LFSGSEGLKFGARIFSGDRDLVQLHPAPRKSVYHSGDPS- 126
 Db 169 RKIDENLLVLRSGELKIDFGSGALLKDTVTDFDGTGVYSPPEWIRYHRYHGRSATV 228
 QY 127 ---GRRILDSYCETWTEAPATGQASSLLAGRLLEQFAA 164
 Db 229 WSLGVLLYDMVCGDIPFE-----QDEILRGRLIFRRRV 263

RESULT 14
 YCDS ECOLI
 ID YCDS ECOLI STANDARD; PRT; 807 AA.
 AC P75907;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ycds precursor.
 GN YCDS OR B1024 OR 21526 OR EGS1270.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MGL655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Iinuma T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155 (1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;

```

RX MEDLINE=20163068; PubMed=10697960;
RA Seki N., Ueki N., Yano K., Saito T., Masuho Y., Muramatsu M.-A.;
RT "CDNA cloning of a novel human gene NAKAP95, neighbor of A-kinase
RN anchoring protein 95 (AKAP95) on chromosome 19p13.11-p13.12 region.";
RC J. Hum. Genet. 45:31-37(2000).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=2022332; PubMed=10761695;
RX Orstavik S., Eide T., Collas P., Han I.O., Tasken K., Kieff E.,
RA Jahnsen T., Skalhegg B.S.;
RT "Identification, cloning and characterization of a novel nuclear
RN protein, HA95, homologous to A-kinase anchoring protein 95.";
RT Biol. Cell 92:27-37(2000).
[3]
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta, and T-cell lymphoma;
RX MEDLINE=20347256; PubMed=10748171;
RA Westberg C., Yang J.-P., Tang H., Reddy T.R., Wong-Staal F.;
RT "A novel shuttle protein binds to RNA helicase A and activates the
RN retroviral constitutive transport element.";
RT J. Biol. Chem. 275:21396-21401(2000).
[4]
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Beak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RN human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[5]
RN SEQUENCE OF 1-358 FROM N.A., AND SUBCELLULAR LOCATION.
RP TISSUE=Fetal brain;
RX MEDLINE=99068504; PubMed=9883615;
RA Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.-A.;
RT "Selection system for genes encoding nuclear-targeted proteins.";
RN Nat. Biotechnol. 16:1338-1342(1998).
CC -!- FUNCTION: Could play a role in constitutive transport element
CC (CTE)-mediated gene expression. Does not seem to be implicated in
CC the binding of regulatory subunit II of PKA.
CC -!- SUBUNIT: Binds to the C-terminal of RNA helicase A.
CC -!- SUBCELLULAR LOCATION: Nuclear at steady state but shuttles between
CC the nucleus and cytoplasm.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB025905; BAA85003.1; -
DR EMBL; AJ243467; CAB65092.1; -
DR EMBL; AF199414; AAF86048.1; -
DR EMBL; BC000713; AAH00713.1; -
DR EMBL; AB015332; BAA34791.1; ALT_INIT.
DR GO; GO:0005634; C:nucleus; TAS.

```

```

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[4]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RN O157:H7 and genomic comparison with a laboratory strain K-12.";
RT DNA Res. 8:11-22(2001).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: STRONG, TO Y.PESTIS HEMIN-BINDING PROTEIN HEMH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000204; AAC74109.1; -
DR EMBL; D90739; BAA35806.1; -
DR EMBL; D90740; BAA35809.1; -
DR EMBL; AE005302; AAG55642.1; -
DR EMBL; AP002554; BAB34693.1; -
DR PIR; F64844; F64844.
DR PIR; F90787; F90787.
DR Ecogen; EG13865; ycds.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
KW Hypothetical protein; Outer membrane; Signal; Complete proteome.
FT SIGNAL 26
FT CHAIN 27 807
FT SEQUENCE 807 AA; 92307 MW; B20067C3D41723FD CRC64;
SQ
Query Match 7.6%; Score 73.5; DB 1; Length 807;
Best Local Similarity 26.1%; Pred. No. 28;
Matches 42; Conservative 16; Mismatches 60; Indels 43; Gaps 9;
QY 2 THQDFQ---LVHLVALNSPQ-PCGMGIRGADPQCFQQAARAGLAGTFRFLSSRLDQL 57
Db 481 TLQEQQAQVLTHTDVEREPQDGVVLRKAVDVHNLAEELRIAGSTG----- 527
QY 58 YSIVRRADRTGVFVNLRLDEVLPF-----SWE--ALFSGSEGQLKPGARFSPDGRDVLQH 111
Db 528 --IDAEGPDGKHVDLTTIYVSPPLKDNWRGFGAGFYADGQFSEKGIV---RDWLAG 581
QY 112 PAWPKSVW-----HGSDPSGRRLLTDSY--CETWR 139
Db 582 VEWRSRNWLEAEVAERVFNHEHXP-CARLSGWYDFNDNWR 621
RESULT 15
NA95 HUMAN
ID NA95 HUMAN STANDARD; PRT: 646 AA.
AC Q5UUX6; Q94792; Q96J58; Q9NRQ0; Q9UGM0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neighbor of A-kinase anchoring protein 95 (Homologous to AKAP95
DE protein) (HA95) (Helicase A-binding protein 95) (HAP95) (HRIHFB2018).
GN NAKAP95.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=Fetal brain;

```

DR GO; GO:0017151; F:DEAD/H-box RNA helicase binding; TAS.
DR InterPro; IPR007071; AKAP95.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF04988; AKAP95; 1.
DR SMART; SM00355; Znf_C2H2; 1.
KW Nuclear protein; Zinc-finger.
FT DOMAIN 274 279 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 362 364 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN_FING 391 413 C2H2-TYPE.
FT ZN_FING 484 507 C2H2-TYPE.
FT DOMAIN 41 52 TYR/GLY-RICH.
FT DOMAIN 602 608 PRO-RICH.
FT DOMAIN 589 597 PRO-RICH.
FT CONFLICT 100 100 D->N (IN REF. 3).
FT CONFLICT 189 189 S->N (IN REF. 3).
FT CONFLICT 351 358 ALTODEN->EFSWGAWC (IN REF. 4).
FT CONFLICT 458 458 Q->H (IN REF. 1).
FT CONFLICT 554 596 PA->RRRRRRRAVETTRRAKQGFRRACRRS
FT CONFLICT 610 640 LPPQSQP (IN REF. 3).
FT CONFLICT 610 640 EEKGAVPLIGGALORIGIPGLDVEDDEE -> GGGGGR
FT CONFLICT 610 640 RGPCMGRGNARSASRASTWTETKK (IN REF. 3).
SQ SEQUENCE 646 AA; 71640 MW; CEC3658F25996BA2 CRC64;
Query Match 7.5%; Score 72.5; DB 1; Length 646;
Best Local Similarity 26.2%; Pred.No. 27;
Matches 39; Conservative 11; Mismatches 68; Indels 31; Gaps 8;
QY 3 HODPOLVHLVALNSPPQGMGRGIRGADFOCFQAAAGLAGTFFRAFLSSRLQDLYSIVR 62
DB 96 NORLDMVPHLE--TDMVGGVYSGGERVDSYESCDSE-----RAVLSER--DLY---- 140
QY 63 RADRTGVVNVLRDEVLPFSWEALFSGSEGLXPGRARIFSPGRDVLQHPMPWPKSVWHG 122
DB 141 ---RSGDYDSELD-----PEMEWAY---EGQYDAYEDQFRMEGNDTF----GPRAQGNAR 185
QY 123 SDPSGRRLTDSYCETWRTTAPATGQASS 151
DB 186 DARGRPWASGYGRWE-DPMGARGQMS 213

Search completed: March 26, 2004, 13:36:34
Job time : 12.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:32:37 ; Search time 32.444 Seconds
(without alignments)
1789.377 Million cell updates/sec

Title: US-09-938-391-4

Perfect score: 966

Sequence: 1 HTHQDFQLVHLVALNSPOP.....CRHAFVVLGIENSVMTPSK 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	835	85.4	187	Q8WX15	Q8WX15 homo sapien
2	835	85.4	261	Q8NG19	Q8NG19 homo sapien
3	835	86.4	816	Q8N434	Q8N434 mus musculus
4	831	86.0	1140	Q61434	Q61434 mus musculus
5	813	84.2	226	Q9QZD2	Q9QZD2 rattus norv
6	754	78.1	171	Q9WUW5	Q9WUW5 rattus norv
7	750	77.6	1344	Q93419	Q93419 gallus gall
8	726	75.2	160	Q9CRT2	Q9CRT2 mus musculus
9	694	71.8	1315	Q8QRL9	Q8QRL9 xenopus lae
10	691	71.5	1307	Q8JFF7	Q8JFF7 xenopus lae
11	631.5	65.4	361	Q8AWC6	Q8AWC6 brachydanio
12	551	57.0	1367	Q35206	Q35206 mus musculus
13	550	56.9	1367	Q1EQD9	Q1EQD9 mus musculus
14	527	54.6	1388	Q9Y4W4	Q9Y4W4 homo sapien
15	457.5	47.4	950	Q86SC8	Q86SC8 ciona intes
16	456	47.2	102	Q96T70	Q96T70 homo sapien

17	434.5	45.0	299	5	Q8MSE3	Q8MSE3 drosophila
18	429.5	44.5	792	5	Q8MT89	Q8MT89 drosophila
19	428.5	44.4	778	5	Q86BH1	Q86BH1 drosophila
20	366	37.9	650	5	Q17866	Q17866 caenorhabdi
21	366	37.9	778	5	Q3U9K6	Q3U9K6 caenorhabdi
22	366	37.9	1117	5	Q3U9X7	Q3U9X7 caenorhabdi
23	366	37.9	1154	5	Q810G6	Q810G6 caenorhabdi
24	101.5	10.5	287	16	Q89138	Q89138 bradyrhizob
25	94.5	9.8	208	16	Q92K28	Q92K28 rhizobium m
26	93.5	9.7	702	2	O52660	O52660 escherichia
27	93.5	9.7	713	2	O52679	O52679 escherichia
28	87	9.0	651	5	Q9VFA9	Q9VFA9 drosophila
29	85.5	8.9	6521	2	Q8RL72	Q8RL72 pseudomonas
30	84.5	8.7	282	16	Q8PJG2	Q8PJG2 xanthomonas
31	84	8.7	539	2	Q814Y1	Q814Y1 streptomyce
32	83.5	8.6	240	16	Q8F863	Q8F863 leptospira
33	83	8.6	557	16	Q9K4L0	Q9K4L0 streptomyce
34	82	8.5	687	4	Q86TP6	Q86TP6 homo sapien
35	82	8.5	891	2	O33468	O33468 pseudomonas
36	82	8.5	1050	16	Q8PAG4	Q8PAG4 xanthomonas
37	82	8.5	1105	4	Q86XT1	Q86XT1 homo sapien
38	82	8.5	1113	4	Q86W03	Q86W03 homo sapien
39	81.5	8.4	190	16	Q9RIZ7	Q9RIZ7 streptomyce
40	81.5	8.4	633	16	Q8XBY5	Q8XBY5 escherichia
41	80.5	8.3	314	16	Q82UU4	Q82UU4 nitrosomona
42	80.5	8.3	569	5	Q9N6F0	Q9N6F0 leishmania
43	79.5	8.2	276	16	Q89P16	Q89P16 bradyrhizob
44	79.5	8.2	346	16	Q9HWN2	Q9HWN2 pseudomonas
45	79.5	8.2	713	16	Q8XEE0	Q8XEE0 escherichia

ALIGNMENTS

RESULT 1

Q8WX15 PRELIMINARY; PRT; 187 AA.
AC Q8WX15;
DT 01-WAR-2002 (TRENBLrel. 20, Created)
DT 01-WAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-WAR-2002 (TRENBLrel. 20, Last annotation update)
DE Collagen XVIII (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX WEDLINE=21409408; PubMed=11517600;
RA Feng Y., Cui L.B., Liu C.X., Ma Q.J.;
RT "Inhibition effect in vitro of purified endostatin expressed in Pichia
RT pastoris.";
RL Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).
DR EMBL; AF416592; AAL37720.1; -;
FT NON_TER
SQ SEQUENCE 187 AA; 20448 MW; 72B1047D85838CD3 CRC64;

Query Match 86.4%; Score 835; DB 4; Length 187;
Best Local Similarity 85.1%; Pred. No. 5.1e-76;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;
QY 1 HTHQDFQLVHLVALNSPOPQGMGRGIRGADFCQQAAGLAGTFRAPLSRLQDLYSI 60
Db 5 HSHRDFQPVHLVHLVALNSPLSGMGRGIRGADFCQQAAGLAGTFRAPLSRLQDLYSI 64
QY 61 VRRADRTGVPVNLDRDLVLPSPWEALFSSGOLKPGARIFSGDGRDLVLPWPAPRKSVM 120
Db 65 VRRADRAVPVNLKDELLFSPWEALFSSGOLKPGARIFSGDGRDLVLPWPAPRKSVM 124
QY 121 HGSPDSRRRTDTSYCETWRTAPAAATQASSLLAGLLEQEAASCRHAFVVLGIENSVM 180
Db 125 HGSPNGRRRTESYCETWRTAPAAATQASSLLAGLLEQEAASCRHAFVVLGIENSVM 184

QY 181 S 181
Db 185 A 185

RESULT 2
Q8NG19
ID Q8NG19 PRELIMINARY; PRT; 261 AA.
AC Q8NG19;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Multi-functional protein MFP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Dou D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF282883; AAM52249.1; -.
SQ SEQUENCE 261 AA; 26745 MW; CA60C920AF3E90E5 CRC64;

Query Match 86.4%; Score 835; DB 4; Length 261;
Best Local Similarity 85.1%; Pred. No. 7.8e-76;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFFAFSSRLQDLYSI 60
Db 79 HSHRDFQVLHLVALNSPLSGGMRGIRGADFCQQAARAVGLAGTFFAFSSRLQDLYSI 138

QY 61 VRRADRTGVPVNLRLDEVLPFSSWEALFSGSEGQKPGARIFSGDGRVLOHPAPWPKSVW 120
Db 139 VRRADRAAIVNLKDELLFPSSWEALFSGSEGQKPGARIFSGDKVLRHPTWPKSVW 198

QY 121 HGSDFSGRLTDSYCEWTWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 199 HGSDFSGRLTDSYCEWTWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 258

QY 181 S 181
Db 259 A 259

RESULT 3
Q8N4S4
ID Q8N4S4 PRELIMINARY; PRT; 816 AA.
AC Q8N4S4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to collagen, type XVIII, alpha 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033715; AA33715.1; -.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD000007; Clg_helix; 1.
KW Collagen.
FT NON_TER
SQ SEQUENCE 816 AA; 82553 MW; 5D539B2946694F86 CRC64;

Query Match 86.4%; Score 835; DB 4; Length 816;
Best Local Similarity 85.1%; Pred. No. 3.3e-75;

Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFFAFSSRLQDLYSI 60
Db 534 HSHRDFQVLHLVALNSPLSGGMRGIRGADFCQQAARAVGLAGTFFAFSSRLQDLYSI 693

QY 61 VRRADRTGVPVNLRLDEVLPFSSWEALFSGSEGQKPGARIFSGDGRVLOHPAPWPKSVW 120
Db 694 VRRADRAAIVNLKDELLFPSSWEALFSGSEGQKPGARIFSGDKVLRHPTWPKSVW 753

QY 121 HGSDFSGRLTDSYCEWTWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 754 HGSDFSGRLTDSYCEWTWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 813

QY 181 S 181
Db 814 A 814

RESULT 4
Q61434
ID Q61434 PRELIMINARY; PRT; 1140 AA.
AC Q61434;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RA Abe N.; Muregaki Y.; Yoshioka H.; Inoue H.; Ninomiya Y.;
RT Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region."
RL Cell. Mol. Biol. Res. 196:576-582(1993).
DR EMBL; D17546; BAA04483.1; -.
DR PIR; B56101; B56101.
DR HSP; P39061; IKOE.
DR MGD; MGI-88449; Coll15a1.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 1.
KW Collagen.
FT NON_TER
SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDFF CRC64;

Query Match 86.0%; Score 831; DB 11; Length 1140;
Best Local Similarity 83.7%; Pred. No. 1.3e-74;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

QY 1 HTHQDFQVLHLVALNSPQGGMRGIRGADFCQQAARAGLAGTFFAFSSRLQDLYSI 60
Db 957 HTHQDFQVLHLVALNSPLSGGMRGIRGADFCQQAARAVGLSGTFFAFSSRLQDLYSI 1016

QY 61 VRRADRTGVPVNLRLDEVLPFSSWEALFSGSEGQKPGARIFSGDGRVLOHPAPWPKSVW 120
Db 1017 VRRADRAAIVNLKDELLFPSSWEALFSGSEGQKPGARIFSGDKVLRHPTWPKSVW 1076

QY 121 HGSDFSGRLTDSYCEWTWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
Db 1077 HGSDFSGRLTDSYCEWTWTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 1136

QY 181 SFSK 184
Db 1137 SFSK 1140

```
RT chronic liver injuries.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AJ236873; CAB44263.1; -.
DR HSP; P39061; IKOE.
FT NON TER 1 171
SQ SEQUENCE 171 AA; 18933 MW; 81BE2BE3FC2C8E72 CRC64;

Query Match 78.1%; Score 754; DB 11; Length 171;
Best Local Similarity 81.9%; Pred. No. 6.5e-68;
Matches 140; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

Qy 9 VHLVALNSPQGMGIRGADFCQFQQAARAGLAGTFRFLSSRLQDLYSIVRRADRTG 68
Db 1 VHLVALNTPLSGGMGIRGADFCQFQQAARAVGLSGTFRFLSSRLQDLYSIVRADRSS 60

Qy 69 VPVNLRLDEVLPFSEALFSGGQKPGARIFSPDGRDVLQHPAWPKSVYHGSDBSGR 128
Db 61 VPVNLRLDEVLPFSEALFSGGQKPGARIFSPDGRDVLQHPAWPKSVYHGSDBSGR 120

Qy 129 RLTDSCYCTWRTEAPATQAASSLLAGRLLEQEAASCRHAFVVLCIENSVM 179
Db 121 RLMEYCYCTWRTEATGTGTGQASSLLSGLLEQKASCHNSVIVLCIENSFM 171

RESULT 7
O93419 PRELIMINARY; PRT; 1344 AA.
ID O93419
AC O93419;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen XVII precursors.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=98411346; PubMed=9738008;
RA Halfter W., Dong S., Schurer B., Cole G.J.;
RT "Collagen XVII is a basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 273:25404-25412 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Halfter W., Dong S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF083440; AAC33294.2; -.
DR HSP; P39061; IKOE.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001791; Laminin_G_lec_gl.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR Collagen; Signal.
FT SIGNAL 1 26 POTENTIAL.
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;

Query Match 77.6%; Score 750; DB 13; Length 1344;
Best Local Similarity 76.1%; Pred. No. 2.3e-66;
Matches 140; Conservative 16; Mismatches 28; Indels 0; Gaps 0;

Qy 1 HTHQDFQVHLVVALNSPQGMGIRGADFCQFQQAARAGLAGTFRFLSSRLQDLYSI 60
Db 1161 HVHQDFQVHLVVALNTPLSGGMGIRGADFCQFQQAARAGLAGTFRFLSSRLQDLYSI 1220
```

```
QY 61 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGQKFGARIFSDGRDVLQHPAMPKRSW 120
DB 1221 VRRADRTAVPIVNLRLDEVLPFSWEALFSGSEGQKFGARIFSDGRDVLQHPAMPKRSW 1280
QY 121 HGSPPSGRRLLTDSYCEWTRTEAPANTGOASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
DB 1281 HGSAPKGRRLPESYCEAWRTDERGTGQASSUSSGKLEQSSCQHPVVLZCIENSFMT 1340
QY 181 SFSK 184
DB 1341 AAKK 1344

RESULT 8
QRCRT2 PRELIMINARY; PRT; 160 AA.
ID Q9CMT2
AC Q9CMT2
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Procollagen, type XVIII, alpha 1 (Fragment).
OS COL18A1.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli P., Barch G.,
RA Blake J., Boffelli D., Bojunga N., Carrinci P., de Bonaldo W.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014292; BAB29249.1; -.
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88451; Coll18a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR GO; GO:0001525; P:angiogenesis; IMP.
FT NON_TER
SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C7375D2 CRC64;

Query Match 75.2%; Score 726; DB 11; Length 160;
Best Local Similarity 83.8%; Pred. No. 4e-65;
Matches 134; Conservative 16; Mismatches 10; Indels 0; Gaps 0;

QY 25 GIRGADFCQCAAGAGTFFRAFLSSRLQDLYSVIRADRTGVPVNLRLDEVLPFSWE 84
DB 1 GIRGADFCQCAAGAGTFFRAFLSSRLQDLYSVIRADRTGVPVNLRLDEVLPFSWE 60
QY 85 ALFSGSEGQKFGARIFSDGRDVLQHPAMPKRSWVHSGSDPSGRLLTDSYCEWTRTEAPA 144
DB 1 SFGSGSQQLQPGARIFSDGRDVLQHPAMPKRSWVHSGSDPSGRLLTDSYCEWTRTEAPA 120
QY 145 ATGOASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFK 184
DB 145 ATGOASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFK 184
```

```
DB 121 ATGOASSLLAGRLLEQEAASCHNSYIVLCIENSFMSTFSK 160
RESULT 9
Q8QHL9 PRELIMINARY; PRT; 1315 AA.
ID Q8QHL9
AC Q8QHL9
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Type XVIII collagen alpha chain.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishino T., Sekimizu K., Natori S., Kubo T.;
RT "Identification and characterization of genes expressed selectively in
RT the regenerating tail of Xenopus laevis tadpole.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047066; BAB84674.1; -.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR03129; TSPN.
DR Pfam; PF01391; Collagen; 7.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; C1g_helix; 1.
DR SMART; SM00210; TSPN; 1.
DR Collagen.
RX SEQUENCE 1315 AA; 134946 MW; 0C56C235DE058365 CRC64;

Query Match 71.8%; Score 694; DB 13; Length 1315;
Best Local Similarity 70.1%; Pred. No. 9.6e-61;
Matches 129; Conservative 18; Mismatches 37; Indels 0; Gaps 0;

QY 1 HTHDFFQLVHLVALNSPQGMRGIRGADFQCFQARAGLAGTFFRAFLSSRLQDLYSI 60
DB 1132 HTHQFNPALHLVALNAPLSGWSKSGVDFQCFQARAGLAGTFFRAFLSSRLQDLYSI 1191
QY 61 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGQKFGARIFSDGRDVLQHPAMPKRSW 120
DB 1192 VRRADRTGVPVNLRLDEVLPFSWEALFSGSEGQKFGARIFSDGRDVLQHPAMPKRSW 1251
QY 121 HGSPPSGRRLLTDSYCEWTRTEAPANTGOASSLLAGRLLEQEAASCRHAFVVLCIENSVM 180
DB 1252 HGSAPKGRRLPESYCEWTRTEAPANTGOASSLLAGRLLEQEAASCRHAFVVLCIENSVM 1311
QY 181 SFSK 184
DB 1312 NNRRK 1315

RESULT 10
Q8JFF7 PRELIMINARY; PRT; 1307 AA.
ID Q8JFF7
AC Q8JFF7
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Type XVIII collagen short variant.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22166979; PubMed=12175494;
RA Elamra H., Peterson J., Pihlajaniemi T., Destree O.;
```


RT "Cloning of three variants of type XVIII collagen and their expression
ET patterns during Xenopus laevis development.";
RL Mech. Dev. 114:109-113(2002).
DR EMBL; AY052763; AAL14257.1; -.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SMO0210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1307 AA; 134153 MW; D53EDFE3DE34976 CRC64;

Query Match 71.5%; Score 691; DB 13; Length 1307;
Best Local Similarity 70.1%; Pred. No. 1.9e-60;
Matches 129; Conservative 18; Mismatches 37; Indels 0; Gaps 0;

QY 1 HTHQDFQLVHLVALNSPQGMGIRGADFCQFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 1124 HTHQDFPALHLVALNAPISGSMKIRGVDVFCFEQARKAGLGTFRFLSSRLQDLYSI 1183

QY 61 VRRADRTGVPVNLRLDEVLPFSWEALFSGSGQLKPGARIFSGDRDVLQHPAWPRKSVW 120
DB 1184 VRRADRSQVIVNLRDEVLDVNDLSLFGSEAGNRSGLRILSPDGKDVTDFTWPKXVW 1243

QY 121 HGSPPSGRLTDSYCYETWTEAPATGQASILLAGRLLEQEAASCRHAFVVLCIENSVM 180
DB 1244 HGSDAKGRRLTESYCYETWTEAPATGQASILLTSGKLEQRPQSNKNPVLICIENSFMT 1303

QY 181 SFSK 184
DB 1304 NNK 1307

RESULT 11
Q8AWC6 PRELIMINARY; PRT; 361 AA.
AC Q8AWC6;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Collagen XVIII (Fragment).
CN COL18A1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Haitek Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
RT "Sequence and embryonic expression of collagen XVIII NCII domain
(endostatin) in the zebrafish.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ494837; CAD38825.1; -.
FT NON_TER 1
SQ SEQUENCE 361 AA; 40222 MW; 3C5A0F8479D26735 CRC64;

Query Match 65.4%; Score 631.5; DB 13; Length 361;
Best Local Similarity 65.2%; Pred. No. 3.7e-55;
Matches 120; Conservative 22; Mismatches 41; Indels 1; Gaps 1;

QY 1 HTHQDFQLVHLVALNSPQGMGIRGADFCQFQQAAGLAGTFRFLSSRLQDLYSI 60
DB 179 HTHTSGP-GLHLIALNSPQGMGIRGADFCQFQQAAGLAGTFRFLSSRLQDLYSI 237

QY 61 VRRADRTGVPVNLRLDEVLPFSWEALFSGSGQLKPGARIFSGDRDVLQHPAWPRKSVW 120
DB 238 VRRSDRETLPIVNLKQDLVFWSSWESLSDSRKNDNAPISYFQGRDVLQHPAWPRKMIW 297

QY 121 HGSPPSGRLTDSYCYETWTEAPATGQASILLAGRLLEQEAASCRHAFVVLCIENSVM 180
DB 1244 HGSDAKGRRLTESYCYETWTEAPATGQASILLTSGKLEQRPQSNKNPVLICIENSFMT 1303

Db 298 HGSDDGRHQRQTDNYCETWRAGDRAVNTGLASSIQAGQLQQTSSSCSSSYALCIENSMT 357
QY 181 SFSK 184
Db 358 QSKK 361

RESULT 12
Q35206 PRELIMINARY; PRT; 1367 AA.
AC Q35206;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Type XV collagen.
DE Type XV collagen.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97480713; PubMed=9339358;
RA Hagb P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
corresponding gene to 4B1-3: Comparison of mouse and human alpha 1
(XV) collagen sequences indicates divergence in the number of small
collagenous domains.";
RT Genomics 45:31-41(1997).
RL EMBL; AF011450; AAC53387.1; -.
DR PDB; 1DY2; 21-JAN-01.
DR MGD; MGI:88449; Coll5a1.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 1.
DR SMART; SMO0282; LamG; 1.
DR SMART; SMO0210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3AEEC CRC64;

Query Match 57.0%; Score 551; DB 11; Length 1367;
Best Local Similarity 59.7%; Pred. No. 2.5e-46;
Matches 105; Conservative 25; Mismatches 42; Indels 4; Gaps 1;

QY 9 VHLVALNSPQGMGIRGADFCQFQQAAGLAGTFRFLSSRLQDLYSIVRRADRTG 68
DB 1196 VHLVALNTPVAGDIR---ADQCQQAARAAGLLTFFRAFLSSHLQDLSTVYKAEFRG 1251

QY 69 VPVNLRLDEVLPFSWEALFSGSGQLKPGARIFSGDRDVLQHPAWPRKSVWHDSPGR 128
DB 1252 LPVNLKQVLFNNWDSIFSGDQGNTHPIYSFDQDVMTDPSWPKYVWHGSPHGV 1311

QY 129 RLTSYCYETWTEAPATGQASILLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 184
DB 1312 RLVDKYCEAWRTTDMVTFASPLSTGKILDQKAYSCANRLIVLICIENSFMTDRK 1367

RESULT 13
Q9EQD9 PRELIMINARY; PRT; 1367 AA.
AC Q9EQD9;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Type XV collagen.

GN COL15A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv;
 RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
 RA Pihlajaniemi T.;
 RT "Cloning of mouse type XV collagen sequences and mapping of the
 RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
 RT (XV) collagen sequences indicates divergence in the number of small
 RT collagenous domains";
 RL Genomics 45:31-41(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv;
 RA Eklund L., Wuona A., Lietard J., Pihlajaniemi T.;
 RT "Structure of the mouse type XV collagen gene, Col15a1, comparison
 RT with the human Col15A1 gene and functional analysis of the promoters
 RT of both genes";
 RL Matrix Biol. 19:489-500(2000).
 DR EMBL; AF261131; AAG27545.1; JOINED.
 DR EMBL; AF261109; AAG27545.1; JOINED.
 DR EMBL; AF261110; AAG27545.1; JOINED.
 DR EMBL; AF261111; AAG27545.1; JOINED.
 DR EMBL; AF261112; AAG27545.1; JOINED.
 DR EMBL; AF261113; AAG27545.1; JOINED.
 DR EMBL; AF261114; AAG27545.1; JOINED.
 DR EMBL; AF261115; AAG27545.1; JOINED.
 DR EMBL; AF261116; AAG27545.1; JOINED.
 DR EMBL; AF261117; AAG27545.1; JOINED.
 DR EMBL; AF261118; AAG27545.1; JOINED.
 DR EMBL; AF261119; AAG27545.1; JOINED.
 DR EMBL; AF261120; AAG27545.1; JOINED.
 DR EMBL; AF261121; AAG27545.1; JOINED.
 DR EMBL; AF261122; AAG27545.1; JOINED.
 DR EMBL; AF261123; AAG27545.1; JOINED.
 DR EMBL; AF261124; AAG27545.1; JOINED.
 DR EMBL; AF261125; AAG27545.1; JOINED.
 DR EMBL; AF261126; AAG27545.1; JOINED.
 DR EMBL; AF261127; AAG27545.1; JOINED.
 DR EMBL; AF261128; AAG27545.1; JOINED.
 DR EMBL; AF261129; AAG27545.1; JOINED.
 DR EMBL; AF261130; AAG27545.1; JOINED.
 DR HSSP; P39061; 1KOE.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; ConA-like lec_gl.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 5.
 DR ProDom; PD000007; Clg helix; 1.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 KW Collagen.
 SQ SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;

Query Match 56.9%; Score 550; DB 11; Length 1367;
 Best Local Similarity 59.7%; Pred. No. 3.2e-46;
 Matches 105; Conservative 25; Mismatches 42; Indels 4; Gaps 1;

QY 9 VHLVVALNSPQGMGIRGAPQCFQQAACAGLACTFAFLSSRLQDLVSVIRADRTG 68
 |||||
 Db 1196 VHLVVALNTVAGDIR-----ADPQCFQQAACAGLACTFAFLSSRLQDLSTVVKARERG 1251
 |||||

QY 69 VPVYVNLRLDEVLPFSWEALFSGSGQLKPGARIFSPDGRDLVQHPAWPKRKSVWHGSDPSGR 128

Db 1252 LPIVNLKQVLFNWDISFSDGQGFNTHPIVSPDGRDVTDPSPQKVVHGSNPYGV 1311
 |||||
 QY 129 RLUTDYSCTWTTEAPATGQASSLLAGRLLEQQAASCRHAFVVLCLSENSVMTSFSK 184
 |||||
 Db 1312 RLVDKXCEAWRTTDMVATGFPASPLSTGKILDKAYSCANRLIVLCLSENSFMTDTRK 1367
 |||||

RESULT 14
 QY4W4 PRELIMINARY; PRT; 1388 AA.
 ID QY4W4
 AC QY4W4;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Type XV collagen.
 OS COL15A1.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 RT exon-intron organization in the 3' region of the corresponding gene";
 RL J. Biol. Chem. 269:4773-4779(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98316357; PubMed=9651385;
 RA Hagg P.M., Wuona A., Lietard J., Kivirikko S., Pihlajaniemi T.;
 RT "Complete exon-intron organization of the human gene for the alpha
 RT chain of type XV collagen (COL15A1) and comparison with the homologous
 RT COL18A1 gene";
 RL J. Biol. Chem. 273:17824-17831(1998).
 DR EMBL; L25280; AAC78500.1; JOINED.
 DR EMBL; AF052956; AAC78500.1; JOINED.
 DR EMBL; AF052957; AAC78500.1; JOINED.
 DR EMBL; AF052958; AAC78500.1; JOINED.
 DR EMBL; AF052959; AAC78500.1; JOINED.
 DR EMBL; AF052960; AAC78500.1; JOINED.
 DR EMBL; AF052961; AAC78500.1; JOINED.
 DR EMBL; AF052962; AAC78500.1; JOINED.
 DR EMBL; AF052963; AAC78500.1; JOINED.
 DR EMBL; AF052964; AAC78500.1; JOINED.
 DR EMBL; AF052965; AAC78500.1; JOINED.
 DR EMBL; AF052966; AAC78500.1; JOINED.
 DR EMBL; AF052967; AAC78500.1; JOINED.
 DR EMBL; AF052968; AAC78500.1; JOINED.
 DR EMBL; AF052969; AAC78500.1; JOINED.
 DR EMBL; AF052970; AAC78500.1; JOINED.
 DR EMBL; AF052971; AAC78500.1; JOINED.
 DR EMBL; AF052972; AAC78500.1; JOINED.
 DR EMBL; AF052973; AAC78500.1; JOINED.
 DR EMBL; AF052974; AAC78500.1; JOINED.
 DR EMBL; AF052975; AAC78500.1; JOINED.
 DR EMBL; L25285; AAC78500.1; JOINED.
 DR EMBL; L25284; AAC78500.1; JOINED.
 DR EMBL; L25283; AAC78500.1; JOINED.
 DR EMBL; L25282; AAC78500.1; JOINED.
 DR EMBL; L25281; AAC78500.1; JOINED.
 DR HSSP; P39061; 1KOE.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; ConA-like lec_gl.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 5.
 DR ProDom; PD000007; Clg helix; 1.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 KW SMART; SM00282; LamG; 1.

DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1388 AA; 141757 MW; 96828E45B847194B CRC64;
Query Match 54.6%; Score 527; DB 4; Length 1388;
Best Local Similarity 56.6%; Pred. No. 6.8e-44;
Matches 99; Conservative 27; Mismatches 45; Indels 4; Gaps 1;
QY 10 LHLVALNSPQGMGIRGADFCQQAARAGLAGTFRAPLSRLQDLYSIVRRADRTGV 69
Db 1218 LHLAALNMFVSGDIR----ADFCQQAARAGLLSTYFAFLSSHLQDLSITVKAERYSL 1273
QY 70 PVNLRDEVLFPSWEALFSGSEGLKPGARIFSPGDRDVLQHPAWPRKSVVHSGDPSGRR 129
Db 1274 PVNLRGQVLFNWDISFGHGQGNMHIPIYSFGRDINTDPSWPKVIWHGSSPHGVR 1333
QY 130 LTDSYCTWRTAPATQOASSLLAGRLLEQPAASCRHAFVVLICIENSVMTSFSK 184
Db 1334 LVNDYCEAWRTADTAVTGLASPLSTGKILDOKAYSCANRLIVLICIENSFMTDARK 1388
RESULT 15
Q86SC8 PRELIMINARY; PRT; 950 AA.
AC Q86SC8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen XVIII homologue.
GN CI-COLLAGEN XVIII.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cionidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RA Hotta K., Takahashi H., Satoh N.;
RT "Further Characterization of Brachyury-Downstream Genes in Ciona
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076898; BAC57521.1; -.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001759; Pentaxin.
DR Pfam; PF01391; Collagen; 8.
DR PROSITE; PS00289; PENTAXIN; 1.
SQ SEQUENCE 950 AA; 94768 MW; 2767ECA158CBA89F CRC64;
Query Match 47.4%; Score 457.5; DB 5; Length 950;
Best Local Similarity 55.2%; Pred. No. 4.2e-37;
Matches 96; Conservative 17; Mismatches 60; Indels 1; Gaps 1;
QY 7 QLVLLHLVALNSPQGMGIRGADFCQQAARAGLAGTFRAPLSRLQDLYSIVRRADR 66
Db 773 QKSLHMIALNFFLRGNTGIVGADARCFQQAARAGLKGTYRAFLSSRDQNVRSIVRRDR 832
QY 67 TGVVNLRLDEVLFPSWEALFSGSEGLKPGARIFSPGDRDVLQHPAWPRKSVVHSGDPS 126
Db 833 RNVPVNIIRGEQLFSSWEELFR-TEGRMDNPNMIYSFENRQVSTDRWPKVFWEGSYTD 891
QY 127 GRLTDSYCTWRTAPATQOASSLLAGRLLEQPAASCRHAFVVLICIENSVMY 180
Db 892 GRLNPMHYCASWYTDHKAVTGCASPLSTRELLAQKPYSCSGEVVLGVENSTRT 945

Search completed: March 26, 2004, 13:38:00
Job time : 34.4444 secs